```
OM nucleic - nucleic search, using sw model

Run on:          September 11, 2004, 10:06:53 ; Search time 14891.6 Seconds
                                            (without alignments)
                                            7040.653 Million cell updates/sec

Title:           US-09-939-408A-12
Perfect score:   2419
Sequence:        1 ctcctgcctcactctaaccc..........gagttcaaaaaaaaaaaaa 2419

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:      6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
                 15:  em_ba:*
                 16:  em_fun:*
                 17:  em_hum:*
                 18:  em_in:*
                 19:  em_mu:*
                 20:  em_om:*
                 21:  em_or:*
                 22:  em_ov:*
                 23:  em_pat:*
                 24:  em_ph:*
                 25:  em_pl:*
                 26:  em_ro:*
                 27:  em_sts:*
                 28:  em_un:*
                 29:  em_vi:*
                 30:  em_htg_hum:*
                 31:  em_htg_inv:*
                 32:  em_htg_other:*
                 33:  em_htg_mus:*
                 34:  em_htg_pln:*
                 35:  em_htg_rod:*
                 36:  em_htg_mam:*
                 37:  em_htg_vrt:*
                 38:  em_sy:*
                 39:  em_htgo_hum:*
                 40:  em_htgo_mus:*
                 41:  em_htgo_other:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2418.6	100.0	2419	6	AR199602	AR199602 Sequence
	2	2418.6	100.0	2419	6	AX366860	AX366860 Sequence
	3	1826.6	75.5	2741	6	AX366876	AX366876 Sequence
	4	1629	67.3	2475	6	AR199606	AR199606 Sequence
	5	1629	67.3	2475	6	AX366868	AX366868 Sequence
	6	1196.6	49.5	2151	6	E01759	E01759 DNA encodin
	7	1196.6	49.5	2151	6	AR205032	AR205032 Sequence
	8	1196.6	49.5	2151	6	AR282639	AR282639 Sequence
	9	1196.6	49.5	2151	6	AX082544	AX082544 Sequence
	10	1196.6	49.5	2151	6	AX370659	AX370659 Sequence
	11	1196.6	49.5	2315	8	RTPALGNA	X12702 R.toruloide
	12	1196.6	49.5	2331	6	E01783	E01783 DNA sequenc
	13	1196.6	49.5	2331	6	E01784	E01784 DNA sequenc
	14	1196.6	49.5	2331	6	E01785	E01785 DNA sequenc
	15	1196.6	49.5	2439	6	AR199605	AR199605 Sequence
	16	1196.6	49.5	2439	6	AX366866	AX366866 Sequence
	17	1195	49.4	2151	6	E01543	E01543 L-phenylala
	18	1195	49.4	2151	6	AR205031	AR205031 Sequence
	19	1195	49.4	2151	6	AR282638	AR282638 Sequence
	20	1195	49.4	2151	6	AX082542	AX082542 Sequence
	21	1195	49.4	2151	6	AX370657	AX370657 Sequence
	22	1188.8	49.1	2061	6	AR282658	AR282658 Sequence
	23	1188.8	49.1	2061	6	AX370681	AX370681 Sequence
	24	1136.6	47.0	2787	6	AR199604	AR199604 Sequence
	25	1136.6	47.0	2787	6	AX366864	AX366864 Sequence
	26	802.4	33.2	2986	8	RHDPAL	M18261 R. toruloid
	27	802.4	33.2	3340	8	RTPAL2	X51513 Rhodosporid
	28	747.2	30.9	3586	8	RRPAL	X13094 Rhodotorula
	29	402	16.6	2124	6	AX653389	AX653389 Sequence
	30	402	16.6	2124	6	AX660330	AX660330 Sequence
	31	402	16.6	2510	8	AK100346	AK100346 Oryza sat
	32	401.4	16.6	2016	8	AY224546	AY224546 Oryza sat
	33	399.8	16.5	2488	8	MZEPAL	L77912 Zea mays ph
	34	399.2	16.5	2126	6	AX660228	AX660228 Sequence
c	35	393.8	16.3	139226	8	OSJN00274	AL731629 Oryza sat
c	36	393.8	16.3	139488	8	OSJN00262	AL731616 Oryza sat
c	37	388.8	16.1	170186	2	AC135419	AC135419 Oryza sat
c	38	388.8	16.1	177374	8	AC135429	AC135429 Oryza sat
	39	385.8	15.9	2504	8	AK102817	AK102817 Oryza sat
	40	384.2	15.9	2211	6	AX660248	AX660248 Sequence
	41	384.2	15.9	2280	6	AX654420	AX654420 Sequence
	42	381.4	15.8	2282	6	AX660254	AX660254 Sequence
	43	379.8	15.7	2594	8	AK120127	AK120127 Oryza sat
	44	379.4	15.7	2157	6	AX654419	AX654419 Sequence
	45	377.8	15.6	2108	6	AX660252	AX660252 Sequence

ALIGNMENTS

```
RESULT 1
AR199602
LOCUS       AR199602                2419 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 12 from patent US 6355468.
ACCESSION   AR199602
VERSION     AR199602.1  GI:20249676
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2419)
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: US 6355468-A 12 12-MAR-2002;
```

```
FEATURES             Location/Qualifiers
     source          1. .2419
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 2418.6;  DB 6;  Length 2419;
  Best Local Similarity   100.0%;  Pred. No. 1e-284;
  Matches 2418;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60

Qy         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120

Qy        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180

Qy        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240

Qy        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300

Qy        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360

Qy        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420

Qy        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480

Qy        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGCCGACGTCCGTCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540

Qy        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600

Qy        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660

Qy        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720

Qy        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780

Qy        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840

Qy        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900

Qy        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960

Qy        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020

Qy       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080

Qy       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140

Qy       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200

Qy       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260

Qy       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320

Qy       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380

Qy       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440

Qy       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500

Qy       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560

Qy       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620

Qy       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680

Qy       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740

Qy       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800

Qy       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860

Qy       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920

Qy       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980

Qy       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040
```

```
Qy       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100

Qy       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160

Qy       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220

Qy       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280

Qy       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340

Qy       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400

Qy       2401 AGTTCAAAAAAAAAAAAAA 2419
              |||||||||||||||||||
Db       2401 AGTTCAAAAAAAAAAAAAA 2419
```

```
RESULT 2
AX366860
LOCUS       AX366860                2419 bp    DNA     linear   PAT 15-FEB-2002
DEFINITION  Sequence 12 from Patent WO0208402.
ACCESSION   AX366860
VERSION     AX366860.1  GI:18698149
KEYWORDS    .
SOURCE      Rhodotorula graminis
  ORGANISM  Rhodotorula graminis
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
            Rhodotorula.
REFERENCE   1
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: WO 0208402-A 12 31-JAN-2002;
            PCBU Services, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2419
                     /organism="Rhodotorula graminis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:29898"
     CDS             37..2199
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD23828.1"
                     /db_xref="GI:18698150"
                     /db_xref="REMTREMBL:CAD23828"
                     /translation="MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGH
                     AAHQSQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDK
                     SVDFLKAQLQNSVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGR
                     GLENTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISAS
                     GDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVN
                     GTAVSASMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEV
                     ARNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTL
                     SLENNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLN
                     AAMNRGLPSCLAAEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVN
                     SLALISARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLG
                     TGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLT
                     AVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQAR
                     RGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA"
     unsure          493..495
                     /note="Other information: Xaa = Val or Ala"
```

```
ORIGIN

  Query Match            100.0%;  Score 2418.6;  DB 6;  Length 2419;
  Best Local Similarity  100.0%;  Pred. No. 1e-284;
  Matches 2419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60

Qy         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120

Qy        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180

Qy        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240

Qy        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300

Qy        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360

Qy        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420

Qy        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480

Qy        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540

Qy        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600

Qy        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660

Qy        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720

Qy        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780

Qy        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840

Qy        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900

Qy        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960

Qy        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020
```

```
Qy       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080

Qy       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140

Qy       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200

Qy       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260

Qy       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320

Qy       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380

Qy       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440

Qy       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500

Qy       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560

Qy       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620

Qy       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680

Qy       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740

Qy       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800

Qy       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860

Qy       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920

Qy       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980

Qy       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040

Qy       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100

Qy       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160

Qy       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220

Qy       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280

Qy       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340

Qy       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400

Qy       2401 AGTTCAAAAAAAAAAAAAA 2419
              |||||||||||||||||||
Db       2401 AGTTCAAAAAAAAAAAAAA 2419
```

```
RESULT 3
AX366876
LOCUS       AX366876                2741 bp    DNA     linear   PAT 15-FEB-2002
DEFINITION  Sequence 28 from Patent WO0208402.
ACCESSION   AX366876
VERSION     AX366876.1  GI:18698164
KEYWORDS    .
SOURCE      Rhodotorula graminis
  ORGANISM  Rhodotorula graminis
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
            Rhodotorula.
REFERENCE   1
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: WO 0208402-A 28 31-JAN-2002;
            PCBU Services, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2741
                     /organism="Rhodotorula graminis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:29898"
     exon            1..361
     intron          362..448
     exon            449..880
     intron          881..960
     exon            961..1295
     intron          1296..1364
     exon            1365..1529
     intron          1530..1586
     exon            1587..1748
     intron          1749..1821
     exon            1822..1947
     intron          1948..2007
     exon            2008..2586
     terminator      2587..2589
ORIGIN

  Query Match             75.5%;  Score 1826.6;  DB 6;  Length 2741;
  Best Local Similarity   84.3%;  Pred. No. 7.4e-213;
  Matches 2312;  Conservative    2;  Mismatches    1;  Indels  426;  Gaps    6;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
              |||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCMCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 156
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 120

Qy       157 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 216
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 180

Qy       217 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 276
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 240

Qy       277 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 336
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 300

Qy       337 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCAC- 395
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db       301 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACG 360

Qy       396 ------------------------------------------------------------ 395

Db       361 GTGCGTTCCGAGACGAGAGGCGGAAATCTCGGGATGCCGCAGCGCTGAACGCTGACACTC 420

Qy       396 --------------------------GGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 429
                                       ||||||||||||||||||||||||||||||||||
Db       421 GCTTGGACGGCTGCCGCGGTCTTGCAGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 480

Qy       430 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 489
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 540

Qy       490 TCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 549
             ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TCCGTCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 600

Qy       550 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 609
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 660

Qy       610 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 669
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 720

Qy       670 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 729
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 780

Qy       730 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 789
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 840

Qy       790 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTC--------------------- 828
             |||||||||||||||||||||||||||||||||||||||                     
Db       841 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGGTACGTCGCGAGTCCTGACT 900

Qy       829 -----------------------------------------------------------G 829
                                                                        |
Db       901 GCAGTGAGCTGTTCGAGAGTCTCCCAGTTTGCTGACTGCCCTTTGTTCATGCGATTGCAG 960

Qy       830 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 889
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 1020

Qy       890 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 949
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 1080

Qy       950 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1009
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1140

Qy      1010 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCT 1069
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCT 1200

Qy      1070 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1129
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1260

Qy      1130 GCCAGGACCGCTACCCGCTCCGCACGTCGCCT---------------------------- 1161
             ||||||||||||||||||||||||||||||||                            
Db      1261 GCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGGTTCGTCCCCTCTCTCTCCCCTTCC 1320

Qy      1162 -----------------------------------------CAGTTCCTCGGCCCGCTCG 1180
                                                      |||||||||||||||||||
Db      1321 CTCCGTCCGACCGGCGCGTCGAGACTTACGTTTGCGTATCCAGTTCCTCGGCCCGCTCG 1380

Qy      1181 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1440

Qy      1241 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1500

Qy      1301 CTGTCTCGATTTCGATGGAGAAGAC----------------------------------- 1325
             |||||||||||||||||||||||||                                   
Db      1501 CTGTCTCGATTTCGATGGAGAAGACCAGGTGCGTCTCTCGCTGCCTTCGTACTCCGATCT 1560

Qy      1326 ----------------------CAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1363
                                   ||||||||||||||||||||||||||||||||||||||
Db      1561 TGTGCTGAATGTTCTTCTCCTGCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1620

Qy      1364 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1680

Qy      1424 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1483
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1740

Qy      1484 CTTCGG------------------------------------------------------ 1489
             ||||||                                                      
Db      1741 CTTCGGAGGTGAGCCGTCGACGTTCTCCGCCGTCGCTCGTCCCCTTCAGCGCACCCAGGC 1800

Qy      1490 -------------------AGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1530
                                |||||||||||||||||||||||||||||||||||||||||
Db      1801 TGACTTCCTTTCCCTCTGTAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1860

Qy      1531 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1590
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1861 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1920

Qy      1591 GCCGAGGCCAACGACGTCCTTTCTCTC--------------------------------- 1617
             |||||||||||||||||||||||||||                                 
Db      1921 GCCGAGGCCAACGACGTCCTTTCTCTCGTGCGTTCGTGTCGCAATGAGTCCCGACGCAAT 1980

Qy      1618 ---------------------------CTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 1650
                                        |||||||||||||||||||||||||||||||||
Db      1981 AGCGACTGACTGCGCGATCCTGAGCAGCTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 2040

Qy      1651 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 1710
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2041 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 2100

Qy      1711 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 1770
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2101 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 2160

Qy      1771 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 1830
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2161 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 2220

Qy      1831 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 1890
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 2280
```

```
Qy       1891 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 1950
Db       2281 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 2340

Qy       1951 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2010
Db       2341 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2400

Qy       2011 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2070
Db       2401 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2460

Qy       2071 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2130
Db       2461 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2520

Qy       2131 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2190
Db       2521 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2580

Qy       2191 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2250
Db       2581 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2640

Qy       2251 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGCCTTACTCTCTCG 2310
Db       2641 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGTCTTACTCTCTCG 2700

Qy       2311 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2351
Db       2701 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2741


RESULT 4
AR199606
LOCUS       AR199606                2475 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 20 from patent US 6355468.
ACCESSION   AR199606
VERSION     AR199606.1  GI:20249680
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2475)
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: US 6355468-A 20 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .2475
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             67.3%;  Score 1629;  DB 6;  Length 2475;
  Best Local Similarity   78.0%;  Pred. No. 7.5e-189;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
```

```
Qy   1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407
Db   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy   1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467
Db   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy   1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527
Db   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy   1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587
Db   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy   1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647
Db   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy   1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707
Db   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy   1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC-----------AACGCA 1755
Db   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740

Qy   1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815
Db   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy   1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875
Db   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy   1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933
Db   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy   1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993
Db   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy   1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053
Db   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy   2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113
Db   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy   2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173
Db   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy   2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232
Db   2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy   2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292
Db   2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy   2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352
Db   2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy   2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411
Db   2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy   2412 AAAAAAA 2419
Db   2401 NNNANAAA 2408
```

```
RESULT 5
AX366868
LOCUS       AX366868                2475 bp    DNA     linear   PAT 15-FEB-2002
DEFINITION  Sequence 20 from Patent WO0208402.
ACCESSION   AX366868
VERSION     AX366868.1  GI:18698157
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: WO 0208402-A 20 31-JAN-2002;
            PCBU Services, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2475
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Consensus Sequence of SEQ ID NOs: 12, 16, and 18"
ORIGIN

  Query Match           67.3%;  Score 1629;  DB 6;  Length 2475;
  Best Local Similarity  78.0%;  Pred. No. 7.5e-189;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy     37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
Db      1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy     97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154
Db     61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy    155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214
Db    121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy    215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271
Db    181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy    272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330
Db    241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy    331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390
Db    301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy    391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450
Db    361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy    451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510
Db    421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy    511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570
Db    481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy    571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630
Db    541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
```

```
Qy     631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690
Db     601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy     691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750
Db     661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy     751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810
Db     721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy     811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870
Db     781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy     871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930
Db     841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy     931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990
Db     901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy     991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050
Db     961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy    1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110
Db    1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy    1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170
Db    1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy    1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227
Db    1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy    1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287
Db    1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy    1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347
Db    1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy    1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407
Db    1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy    1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467
Db    1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy    1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527
Db    1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy    1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587
Db    1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy    1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647
Db    1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy    1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707
Db    1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy    1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC------------AACGCA 1755
Db    1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740

Qy    1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815
Db    1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy    1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875
Db    1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy    1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933
Db    1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy    1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993
Db    1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy    1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053
Db    1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy    2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113
Db    2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy    2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173
Db    2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy    2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232
Db    2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy    2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292
Db    2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNN 2280

Qy    2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352
Db    2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy    2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411
Db    2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy    2412 AAAAAAAA 2419
Db    2401 NNNANAAA 2408
```

RESULT 6
E01759

```
LOCUS       E01759                  2151 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  DNA encoding L-phenylanine anmonia-lyase.
ACCESSION   E01759
VERSION     E01759.1  GI:2170012
KEYWORDS    JP 1988291583-A/1.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2151)
  AUTHORS   Fukuhara,N., Yoshino,S., Yamamoto,K., Nakajima,Y., Iwata,T.,
            Watanabe,M., Suzuki,M. and Makiguchi,N.
  TITLE     AMINO ACID SEQUENCE OF L-PHENYLALANINE AMMONIA LYASE, ITS
            STRUCTURAL GENE, NOVEL BASE SEQUENCE CONTAINING SAME, TRANSFORMANT
            PRODUCED THEREWITH AND PRODUCTION OF L-PHENYLALANINE USING SAME
  JOURNAL   Patent: JP 1988291583-A 1 29-NOV-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium toruloides
            PN   JP 1988291583-A/1
```

```
     PD    29-NOV-1988
     PF    22-MAY-1987 JP 1987123950
     PI    FUKUHARA NOBUHIRO, YOSHINO SADAO, YAMAMOTO KAORU, PI
     NAKAJIMA YOSHIYUKI,
     PI    IWATA TOMOYUKI, WATANABE MIDORI, SUZUKI MAKI, PI   MAKIGUCHI
     NOBUYOSHI
     PC    C12N9/88,C12N1/16,C12N1/18,C12N1/20,C12N15/00,C12P13/22, PC
     (C12P13/22,
     PC    C12R1:19),(C12P13/22,C12R1:865);
     CC    strandedness: Double;
     CC    topology: Linear;
     CC    hypothetical: No;
     CC    anti-sense: No;
     CC    *source: strain=IFO 559;
     CC    *source: clone=pSW101;
     FH    Key             Location/Qualifiers
     FH
     FT    CDS             1. .2151
     FT                    /product='L-phenylananine ammonia-lyase' FT
     mat_peptide     1. .2148
     FT                    /product='L-phenylanine ammonia-lyase'.
FEATURES             Location/Qualifiers
     source          1. .2151
                     /organism="Rhodosporidium toruloides"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:5286"
ORIGIN

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 2.5e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
```

```
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 7
AR205032
LOCUS AR205032 2151 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 9 from patent US 6368837.
ACCESSION AR205032
VERSION AR205032.1 GI:21502510
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 2151)
AUTHORS Gatenby,A.A., Sariaslani,S., Tang,X.-S., Qi,W.Wei. and Vannelli,T.
TITLE Bioproduction of para-hydroxycinnamic acid
JOURNAL Patent: US 6368837-A 9 09-APR-2002;
FEATURES Location/Qualifiers
source 1. .2151
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.5e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

```
Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424
```

```
Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 8

AR282639

```
LOCUS       AR282639                2151 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 9 from patent US 6521748.
ACCESSION   AR282639
VERSION     AR282639.1  GI:29719238
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2151)
  AUTHORS   Tang,X.-S.
  TITLE     Polynucleotide encoding a mutant Rhodotorula glutinis tyrosine
            ammonia lyase polypeptide
  JOURNAL   Patent: US 6521748-A 9 18-FEB-2003;
FEATURES             Location/Qualifiers
     source          1..2151
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 2.5e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064
```

```
Qy     1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db     1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy     1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db     1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy     1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db     1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy     1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db     1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy     1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db     1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy     1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db     1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy     1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db     1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy     1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db     1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy     1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db     1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy     1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db     1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy     1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db     1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy     1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db     1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198
Db     2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 9
AX082544
LOCUS       AX082544                2151 bp    DNA     linear   PAT 28-FEB-2001
DEFINITION  Sequence 9 from Patent WO0111071.
ACCESSION   AX082544
VERSION     AX082544.1  GI:13184657
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Tang,X.S., Vannelli,T.M., Qi,W.W., Sariaslani,S. and Gatenby,A.A.
  TITLE     Bioproduction of para-hydroxycinnamic acid
  JOURNAL   Patent: WO 0111071-A 9 15-FEB-2001;
            E.I. DU PONT DE NEMOURS & COMPANY INCORPORATED (US)
FEATURES             Location/Qualifiers
     source          1..2151
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="mutant from Rhodotorula glutinis"
ORIGIN

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 2.5e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy       99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db       45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy      159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db      105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy      219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db      165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy      279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db      225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy      339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db      285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy      399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db      345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy      459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db      405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy      519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db      465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy      579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db      525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy      639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db      585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy      699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
```

```
Db     645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy     759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db     705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy     819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db     765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy     879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db     825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy     939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db     885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy     999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db     945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy    1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db    1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy    1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db    1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy    1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db    1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy    1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db    1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy    1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db    1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy    1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db    1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy    1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db    1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy    1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db    1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy    1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db    1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy    1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db    1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy    1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db    1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy    1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db    1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy    1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db    1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy    1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db    1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy    1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db    1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy    1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db    1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy    2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db    1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy    2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db    2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy    2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db    2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy    2184 CAAGATGCTCGCGTA 2198
Db    2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 10
AX370659
LOCUS       AX370659               2151 bp    DNA     linear   PAT 01-MAR-2002
DEFINITION  Sequence 9 from Patent WO0210407.
ACCESSION   AX370659
VERSION     AX370659.1  GI:19168824
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Gatenby,A.A., Sariaslani,F.S., Tang,X.S., Qi,W.W. and Vannelli,T.
  TITLE     Bioproduction of para-hydroxycinnamic acid
  JOURNAL   Patent: WO 0210407-A 9 07-FEB-2002;
            E.I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES             Location/Qualifiers
     source          1..2151
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     CDS             1..2151
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD26694.1"
                     /db_xref="GI:19168825"
                     /db_xref="REMTREMBL:CAD26694"
                     /translation="MAPSLDSISHSFANGVASAKQAVNGASTNLAVAGSHLPTTQVTQ
                     VDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLR
                     SQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSL
                     PLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL
                     SYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSA
                     SMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRK
                     LLEGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQ
                     STTDNPLIDVENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNR
                     GLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALI
                     SARRTTESNDVLSLLLATHLYCVLQAIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTG
                     SNLRDELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNA
                     WKVAAAESAISLTRQVRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDV
                     FLGKQEVTIGSNVSKIYEAIKSGRINNVLLKMLA"
     misc_feature    1..2151
                     /note="mutant from Rhodotorula glutinis"
```

```
ORIGIN

  Query Match            49.5%;  Score 1196.6;  DB 6;  Length 2151;
  Best Local Similarity  74.0%;  Pred. No. 2.5e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135
```

```
Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150

RESULT 11
RTPALGNA
LOCUS       RTPALGNA                2315 bp    mRNA    linear   PLN 23-MAR-1993
DEFINITION  R.toruloides mRNA for L-phenylalanine ammonia-lyase (PAL).
ACCESSION   X12702
VERSION     X12702.1  GI:288350
KEYWORDS    L-phenylalanine ammonia-lyase; PAL gene.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2315)
  AUTHORS   Unknown.
  JOURNAL   Unpublished
COMMENT     (Patented) EP_0_260_919_A1.
FEATURES             Location/Qualifiers
     source          1..2315
                     /organism="Rhodosporidium toruloides"
                     /mol_type="mRNA"
                     /strain="IFO 559"
                     /db_xref="taxon:5286"
                     /clone="pSW13"
     gene            1..2315
                     /gene="PAL"
     CDS             1..2151
                     /gene="PAL"
                     /codon_start=1
                     /product="L-phenylalanine ammonia-lyase"
                     /protein_id="CAA31209.1"
                     /db_xref="GI:288351"
                     /db_xref="GOA:P11544"
                     /db_xref="SWISS-PROT:P11544"
                     /translation="MAPSLDSISHSFANGVASAKQAVNGASTNLAVAGSHLPTTQVTQ
                     VDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLR
                     SQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSL
                     PLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL
                     SYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSA
                     SMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRK
                     LLEGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQ
                     STTDNPLIDVENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNR
                     GLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALI
                     SARRTTESNDVLSLLLATHLYCVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTG
                     SNLRDELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNA
                     WKVAAAESAISLTRQVRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDV
                     FLGKQEVTIGSNVSKIYEAIKSGRINNVLLKMLA"
     polyA_site      2315
                     /gene="PAL"
ORIGIN

  Query Match             49.5%;  Score 1196.6;  DB 8;  Length 2315;
  Best Local Similarity   74.0%;  Pred. No. 2.4e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364
```

```
Qy     1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db     1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy     1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db     1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy     1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db     1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy     1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db     1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy     1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db     1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy     1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db     1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy     1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db     1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198
Db     2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 12
E01783
LOCUS       E01783                  2331 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for L-phenylalanine ammonialyase.
ACCESSION   E01783
VERSION     E01783.1  GI:2170036
KEYWORDS    JP 1988317086-A/1.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2331)
  AUTHORS   Fukuhara,N., Yoshino,S., Watanabe,M., Nakajima,Y. and Makiguchi,N.
  TITLE     RECOMBINANT PLASMID FOR MANIFESTING L-PHENYLALANINE-AMMONIA-LYASE
            AND TRANSFORMANT CONTAINING SAID PLASMID
  JOURNAL   Patent: JP 1988317086-A 1 26-DEC-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium toruloides
            PN   JP 1988317086-A/1
            PD   26-DEC-1988
            PF   18-JUN-1987 JP 1987152357
            PI   FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
            NAKAJIMA YOSHIYUKI,
            PI   MAKIGUCHI NOBUYOSHI
            PC   C12N15/00,C12N1/20//C12N9/88,(C12N1/20,C12R1:01),(C12N9/88, PC
              C12R1:01);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=IFO 559 (ATCC 10788);
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1. .2151
            FT                   /product='L-phenylalanine ammonialyase' FT
            3'UTR           2152. .2331.
FEATURES             Location/Qualifiers
     source          1. .2331
                     /organism="Rhodosporidium toruloides"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5286"
ORIGIN

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2331;
  Best Local Similarity   74.0%;  Pred. No. 2.4e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy       99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db       45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy      159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db      105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy      219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db      165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy      279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db      225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy      339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db      285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy      399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db      345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy      459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db      405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy      519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db      465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy      579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db      525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy      639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db      585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy      699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
```

```
Db    645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy    759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db    705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy    819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db    765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy    879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db    825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy    939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db    885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy    999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db    945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy   1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db   1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy   1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db   1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy   1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db   1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy   1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db   1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy   1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db   1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy   1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db   1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy   1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db   1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy   1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db   1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy   1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db   1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy   1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db   1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy   1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db   1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy   1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db   1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy   1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db   1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy   1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db   1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy   1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db   1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy   1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db   1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy   2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db   1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy   2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db   2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy   2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db   2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy   2184 CAAGATGCTCGCGTA 2198
Db   2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 13
E01784
LOCUS       E01784                  2331 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for L-phenylalanine ammonialyase.
ACCESSION   E01784
VERSION     E01784.1  GI:2170037
KEYWORDS    JP 1988317087-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2331)
  AUTHORS   Fukuhara,N., Yoshino,S., Watanabe,M., Nakajima,Y. and Makiguchi,N.
  TITLE     CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF
            ADVENTITIOUS GENE PRODUCT USING SAID CONTROL
  JOURNAL   Patent: JP 1988317087-A 1 26-DEC-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium
            PN   JP 1988317087-A/1
            PD   26-DEC-1988
            PF   18-JUN-1987 JP 1987152358
            PI   FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
            NAKAJIMA YOSHIYUKI,
            PI   MAKIGUCHI NOBUYOSHI
            PC   C12N15/00,C12N9/88,(C12N9/88,C12R1:01);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=IFO 559 (ATCC 10788);
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1. .2151
            FT                   /product='L-phenylalanine ammonialyase' FT
            3'UTR           2152. .2331.
FEATURES             Location/Qualifiers
     source          1. .2331
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match            49.5%;  Score 1196.6;  DB 6;  Length 2331;
  Best Local Similarity  74.0%;  Pred. No. 2.4e-136;
```

```
Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 14
E01785
LOCUS       E01785                  2331 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for L-plenylalanine ammonialyase.
ACCESSION   E01785
VERSION     E01785.1  GI:2170038
KEYWORDS    JP 1988317088-A/1.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2331)
  AUTHORS   Fukuhara,N., Yoshino,S., Watanabe,M., Yamamoto,K., Suzuki,M. and
            Nakajima,Y.
  TITLE     CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF
            ADVENTITIOUS GENE PRODUCT USING SAID CONTROL
  JOURNAL   Patent: JP 1988317088-A 1 26-DEC-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium toruloides
            PN   JP 1988317088-A/1
            PD   26-DEC-1988
            PF   18-JUN-1987 JP 1987152359
            PI   FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
            YAMAMOTO KAORU,
            PI   SUZUKI MAKI, NAKAJIMA YOSHIYUKI
            PC   C12N15/00,C12N9/88,(C12N9/88,C12R1:01);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=IFO 559 (ATCC 10788);
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1. .2151
            FT                   /product='L-phenylalanine ammonialyase' FT
            3'UTR           2152. .2331.
FEATURES             Location/Qualifiers
     source          1. .2331
                     /organism="Rhodosporidium toruloides"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5286"
ORIGIN

  Query Match            49.5%;  Score 1196.6;  DB 6;  Length 2331;
  Best Local Similarity  74.0%;  Pred. No. 2.4e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
```

```
Db     1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy     1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db     1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy     1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db     1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy     1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db     1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy     1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db     1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198
Db     2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 15
AR199605
LOCUS       AR199605                2439 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 18 from patent US 6355468.
ACCESSION   AR199605
VERSION     AR199605.1  GI:20249679
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2439)
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: US 6355468-A 18 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..2439
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2439;
  Best Local Similarity   74.0%;  Pred. No. 2.4e-136;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy       99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db       45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy      159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db      105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy      219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db      165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy      279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db      225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy      339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db      285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy      399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db      345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy      459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db      405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy      519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db      465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy      579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db      525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy      639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db      585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy      699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db      645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy      759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db      705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy      819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db      765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy      879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db      825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy      939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db      885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy      999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db      945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy     1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db     1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy     1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db     1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy     1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
```

```
Db     1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy     1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db     1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy     1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db     1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy     1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db     1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy     1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db     1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy     1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db     1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy     1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db     1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy     1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db     1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy     1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db     1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy     1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db     1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy     1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db     1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198
Db     2136 CAAGATGCTCGCTTA 2150
```

Search completed: September 11, 2004, 23:12:38
Job time : 14904.6 secs

GenCore version 5.1.6


```
OM nucleic - nucleic search, using sw model

Run on:          September 11, 2004, 10:03:28 ; Search time 1334.06 Seconds
                                           (without alignments)
                                           7703.105 Million cell updates/sec

Title:           US-09-939-408A-12
Perfect score:   2419
Sequence:        1 ctcctgcctcactctaaccc..........gagttcaaaaaaaaaaaaaa 2419

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_29Jan04:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002s:*
                 7:  geneseqn2003as:*
                 8:  geneseqn2003bs:*
                 9:  geneseqn2003cs:*
                 10:  geneseqn2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2418.6	100.0	2419	6	ABA95240	Aba95240 R. gramin
2	2418.6	100.0	2419	6	AAD45812	Aad45812 Rhodotoru
3	1826.6	75.5	2741	6	ABA95251	Aba95251 R. gramin
4	1826.6	75.5	2741	6	AAD45823	Aad45823 Rhodotoru
5	1704	70.4	2163	6	AAD45824	Aad45824 Yeast phe
6	1629	67.3	2475	6	ABA95244	Aba95244 Rhodotoru
7	1629	67.3	2475	6	AAD45816	Aad45816 Phenylala
8	1198.2	49.5	2151	6	AAD33097	Aad33097 R. glutin
9	1196.6	49.5	2151	6	AAD33076	Aad33076 Rhodotoru
10	1196.6	49.5	2151	6	AAD33100	Aad33100 R. glutin
11	1196.6	49.5	2151	6	AAD26924	Aad26924 Rhodotoru
12	1196.6	49.5	2151	7	ABS56543	Abs56543 R. glutin
13	1196.6	49.5	2151	9	ADE36055	Ade36055 Rhodospor
14	1196.6	49.5	2331	1	AAN81101	Aan81101 Sequence
15	1196.6	49.5	2331	1	AAN81116	Aan81116 L-phenyla
16	1196.6	49.5	2439	6	ABA95243	Aba95243 R. torulo
17	1196.6	49.5	2439	6	AAD45815	Aad45815 Rhodotoru
18	1195	49.4	2151	6	AAD33152	Aad33152 R. glutin
19	1195	49.4	2151	6	AAD33075	Aad33075 Rhodotoru
20	1195	49.4	2151	6	AAD26923	Aad26923 Rhodotoru
21	1195	49.4	2151	7	ABS56530	Abs56530 R. glutin
22	1195	49.4	2151	9	ADE36053	Ade36053 Rhodospor
23	1195	49.4	2151	9	ADE36061	Ade36061 Rhodospor
24	1195	49.4	2151	9	ADE36075	Ade36075 Rhodospor
25	1195	49.4	2151	9	ADE36065	Ade36065 Rhodospor
26	1195	49.4	2328	1	AAN81399	Aan81399 Sequence
27	1193.4	49.3	2151	1	AAN91092	Aan91092 Structura
28	1193.4	49.3	2151	6	AAD33099	Aad33099 R. glutin
29	1193.4	49.3	2151	6	AAD33096	Aad33096 R. glutin
30	1193.4	49.3	2151	6	AAD33098	Aad33098 R. glutin
31	1193.4	49.3	2151	9	ADE36059	Ade36059 Rhodospor
32	1193.4	49.3	2151	9	ADE36067	Ade36067 Rhodospor
33	1193.4	49.3	2151	9	ADE36073	Ade36073 Rhodospor
34	1193.4	49.3	2151	9	ADE36057	Ade36057 Rhodospor
35	1191.8	49.3	2151	9	ADE36069	Ade36069 Rhodospor
36	1190.2	49.2	2151	9	ADE36063	Ade36063 Rhodospor
37	1190.2	49.2	2151	9	ADE36071	Ade36071 Rhodospor
38	1188.8	49.1	2061	6	AAD33095	Aad33095 Rhodotoru
39	1173.6	48.5	2085	1	AAN82309	Aan82309 Intron-fr
40	1136.6	47.0	2787	6	ABA95242	Aba95242 R. rubra/
41	1136.6	47.0	2787	6	AAD45814	Aad45814 Rhodotoru
42	784	32.4	2987	1	AAN80456	Aan80456 Sequence
43	630.4	26.1	1218	1	AAN80458	Aan80458 Sequence
44	620.6	25.7	1355	1	AAN80457	Aan80457 Sequence
45	402	16.6	2124	7	ADA69936	Ada69936 Rice gene

ALIGNMENTS

```
RESULT 1
ABA95240
ID   ABA95240 standard; cDNA; 2419 BP.
XX
AC   ABA95240;
XX
DT   10-JUN-2002  (first entry)
XX
DE   R. graminis PAL polypeptide encoding cDNA.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate; ss;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             37. .2199
FT                   /*tag=  a
FT                   /transl_except= (pos: 493. .495, aa: Xaa)
FT                   /note= "Xaa = Val or Ala"
FT                   /product= "phenylalanine ammonia lyase"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   P-PSDB; ABB07690.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Claim 4; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
```

```
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a R. graminis PAL polypeptide encoding cDNA
XX
SQ   Sequence 2419 BP; 416 A; 909 C; 659 G; 434 T; 0 U; 1 Other;

  Query Match            100.0%;  Score 2418.6;  DB 6;  Length 2419;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60

Qy         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120

Qy        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180

Qy        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240

Qy        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300

Qy        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360

Qy        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420

Qy        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480

Qy        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540

Qy        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600

Qy        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660

Qy        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720

Qy        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780

Qy        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840

Qy        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900

Qy        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960

Qy        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020

Qy       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080

Qy       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140

Qy       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200

Qy       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260

Qy       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320

Qy       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380

Qy       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440

Qy       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500

Qy       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560

Qy       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620

Qy       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680

Qy       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740

Qy       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800

Qy       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860

Qy       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920

Qy       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980

Qy       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040

Qy       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100

Qy       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160

Qy       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220

Qy       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280

Qy       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340

Qy       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400

Qy       2401 AGTTCAAAAAAAAAAAAAA 2419
              |||||||||||||||||||
Db       2401 AGTTCAAAAAAAAAAAAAA 2419


RESULT 2
AAD45812
ID   AAD45812 standard; DNA; 2419 BP.
XX
AC   AAD45812;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Rhodotorula graminis PAL DNA #1.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; gene; ds.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             37. .2199
FT                   /*tag=  a
FT                   /product= "PAL protein"
FT                   /EC_number= "4.3.1.5"
FT                   /transl_except= (pos:493. .495, aa:Xaa)
FT                   /note= "Xaa corresponds to Val or Ala"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   P-PSDB; AAE27938.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 13; Page 27-29; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is R. graminis PAL DNA
XX
SQ   Sequence 2419 BP; 416 A; 909 C; 659 G; 434 T; 0 U; 1 Other;

  Query Match             100.0%;  Score 2418.6;  DB 6;  Length 2419;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60

Qy         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120

Qy        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180

Qy        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240

Qy        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300

Qy        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360

Qy        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420

Qy        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480

Qy        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540

Qy        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600

Qy        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660

Qy        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720

Qy        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780

Qy        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840

Qy        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900

Qy        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960

Qy        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020

Qy       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080

Qy       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140

Qy       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200

Qy       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260

Qy       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320

Qy       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380

Qy       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440

Qy       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500

Qy       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560

Qy       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620

Qy       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680

Qy       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740

Qy       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800

Qy       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860

Qy       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920

Qy       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980

Qy       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040

Qy       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100

Qy       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160

Qy       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220

Qy       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280

Qy       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340

Qy       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400

Qy       2401 AGTTCAAAAAAAAAAAAAA 2419
              |||||||||||||||||||
Db       2401 AGTTCAAAAAAAAAAAAAA 2419
```

```
RESULT 3
ABA95251
ID   ABA95251 standard; DNA; 2741 BP.
XX
AC   ABA95251;
XX
DT   10-JUN-2002  (first entry)
XX
DE   R. graminis PAL genomic DNA sequence.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate; ds;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2589
FT                   /*tag=  d
FT                   /product= "PAL polypeptide"
FT                   /note= "contains introns"
FT   exon            1. .361
FT                   /*tag=  c
FT                   /number= 1
FT   misc_feature    6
FT                   /*tag=  b
FT                   /note= "PAL DNA of strain ATCC 20804 at this position
FT                   contains the nucleotide C and the PAL DNA of strain KGX
FT                   39 contains the nucleotide A"
FT   intron          362. .448
FT                   /*tag=  d
FT                   /number= 1
```

```
FT   exon            449. .880
FT                   /*tag=  e
FT                   /number= 2
FT   intron          881. .960
FT                   /*tag=  f
FT                   /number= 2
FT   exon            961. .1295
FT                   /*tag=  g
FT                   /number= 3
FT   intron          1296. .1364
FT                   /*tag=  h
FT                   /number= 3
FT   exon            1365. .1529
FT                   /*tag=  i
FT                   /number= 4
FT   intron          1530. .1586
FT                   /*tag=  j
FT                   /number= 4
FT   exon            1587. .1748
FT                   /*tag=  k
FT                   /number= 5
FT   intron          1749. .1821
FT                   /*tag=  l
FT                   /number= 5
FT   exon            1822. .1947
FT                   /*tag=  m
FT                   /number= 6
FT   intron          1948. .2007
FT                   /*tag=  n
FT                   /number= 6
FT   exon            2008. .2586
FT                   /*tag=  o
FT                   /number= 7
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   P-PSDB; ABB07690.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Claim 4; Fig 3, 4; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a R. graminis PAL genomic DNA sequence
XX
SQ   Sequence 2741 BP; 441 A; 1019 C; 756 G; 524 T; 0 U; 1 Other;

  Query Match             75.5%;  Score 1826.6;  DB 6;  Length 2741;
  Best Local Similarity   84.3%;  Pred. No. 1.5e-280;
  Matches 2312;  Conservative    2;  Mismatches    1;  Indels  426;  Gaps    6;
```

```
Qy     37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
          ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCMCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 60

Qy     97 CACGCCGCTCCGACCAAGTCGGCTGCGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CACGCCGCTCCGACCAAGTCGGCTGCGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 120

Qy    157 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 180

Qy    217 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 240

Qy    277 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 300

Qy    337 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCAC- 395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    301 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACG 360

Qy    396 ------------------------------------------------------------ 395

Db    361 GTGCGTTCCGAGACGAGAGGCGGAAATCTCGGGATGCCGCAGCGCTGAACGCTGACACTC 420

Qy    396 --------------------------GGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 429
                                    ||||||||||||||||||||||||||||||||||
Db    421 GCTTGGACGGCTGCCGCGGTCTTGCAGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 480

Qy    430 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 540

Qy    490 TCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 549
          ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCCGTCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 600

Qy    550 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 660

Qy    610 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 720

Qy    670 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 780

Qy    730 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 840

Qy    790 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTC--------------------- 828
          |||||||||||||||||||||||||||||||||||||||                     
Db    841 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGGTACGTCGCGAGTCCTGACT 900

Qy    829 -----------------------------------------------------------G 829
                                                                     |
Db    901 GCAGTGAGCTGTTCGAGAGTCTCCCAGTTTGCTGACTGCCCTTTGTTCATGCGATTGCAG 960

Qy    830 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 1020

Qy    890 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 1080
```

```
Qy     950 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1140

Qy    1010 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCAACATCCGCACGCTCCTTTCCGGCT 1069
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCAACATCCGCACGCTCCTTTCCGGCT 1200

Qy    1070 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1260

Qy    1130 GCCAGGACCGCTACCCGCTCCGCACGTCGCCT---------------------------- 1161
           ||||||||||||||||||||||||||||||||
Db    1261 GCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGGTTCGTCCCCTCTCTCTCCCCTTCC 1320

Qy    1162 -----------------------------------------CAGTTCCTCGGCCCGCTCG 1180
                                                    |||||||||||||||||||
Db    1321 CTCCGTCCGACCGGCGCGTCGAGACTTACGTTTTGCGTATCCAGTTCCTCGGCCCGCTCG 1380

Qy    1181 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1440

Qy    1241 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1500

Qy    1301 CTGTCTCGATTTCGATGGAGAAGAC----------------------------------- 1325
           |||||||||||||||||||||||||
Db    1501 CTGTCTCGATTTCGATGGAGAAGACCAGGTGCGTCTCTCGCTGCCTTCGTACTCCGATCT 1560

Qy    1326 ----------------------CAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1363
                                 ||||||||||||||||||||||||||||||||||||||
Db    1561 TGTGCTGAATGTTCTTCTCCTGCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1620

Qy    1364 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1680

Qy    1424 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1740

Qy    1484 CTTCGG------------------------------------------------------ 1489
           ||||||
Db    1741 CTTCGGAGGTGAGCCGTCGACGTTCTCCGCCGTCGCTCGTCCCCTTCAGCGCACCCAGGC 1800

Qy    1490 -------------------AGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1530
                              |||||||||||||||||||||||||||||||||||||||||
Db    1801 TGACTTCCTTTCCCTCTGTAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1860

Qy    1531 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1920

Qy    1591 GCCGAGGCCAACGACGTCCTTTCTCTC--------------------------------- 1617
           |||||||||||||||||||||||||||
Db    1921 GCCGAGGCCAACGACGTCCTTTCTCTCGTGCGTTCGTGTCGCAATGAGTCCCGACGCAAT 1980

Qy    1618 ---------------------------CTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 1650
                                      |||||||||||||||||||||||||||||||||
Db    1981 AGCGACTGACTGCGCGATCCTGAGCAGCTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 2040

Qy    1651 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 1710
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 2100

Qy    1711 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 1770
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 2160

Qy    1771 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 1830
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 2220

Qy    1831 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 1890
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 2280

Qy    1891 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 1950
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 2340

Qy    1951 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2010
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2400

Qy    2011 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2070
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2460

Qy    2071 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2520

Qy    2131 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2580

Qy    2191 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2640

Qy    2251 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGCCTTACTCTCTCG 2310
           ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    2641 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGTCTTACTCTCTCG 2700

Qy    2311 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2351
           |||||||||||||||||||||||||||||||||||||||||
Db    2701 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2741
```

```
RESULT 4
AAD45823
ID   AAD45823 standard; DNA; 2741 BP.
XX
AC   AAD45823;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Rhodotorula graminis PAL genomic DNA.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; gene; ds.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2589
FT                   /*tag= a
FT                   /product= "PAL protein"
FT                   /EC_number= "4.3.1.5"
FT   exon            1. .361
FT                   /*tag= b
FT   intron          362. .448
FT                   /*tag= c
FT   exon            449. .880
FT                   /*tag= d
FT   intron          881. .960
FT                   /*tag= e
FT   exon            961. .1295
FT                   /*tag= f
FT   intron          1296. .1364
FT                   /*tag= g
```

```
FT   exon            1365. .1529
FT                   /*tag=  h
FT   intron          1530. .1586
FT                   /*tag=  i
FT   exon            1587. .1748
FT                   /*tag=  j
FT   intron          1749. .1821
FT                   /*tag=  k
FT   exon            1822. .1947
FT                   /*tag=  l
FT   intron          1948. .2007
FT                   /*tag=  m
FT   exon            2008. .2589
FT                   /*tag=  n
FT   terminator      2587. .2589
FT                   /*tag=  o
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 4; Page 50-53; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is R. graminis PAL
CC   genomic DNA
XX
SQ   Sequence 2741 BP; 441 A; 1019 C; 756 G; 524 T; 0 U; 1 Other;

  Query Match            75.5%;  Score 1826.6;  DB 6;  Length 2741;
  Best Local Similarity  84.3%;  Pred. No. 1.5e-280;
  Matches 2312;  Conservative    2;  Mismatches    1;  Indels  426;  Gaps    6;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
              |||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCMCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 120

Qy        157 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 180

Qy        217 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 240

Qy        277 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 300

Qy        337 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCAC- 395
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db        301 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACG 360

Qy        396 ------------------------------------------------------------ 395

Db        361 GTGCGTTCCGAGACGAGAGGCGGAAATCTCGGGATGCCGCAGCGCTGAACGCTGACACTC 420

Qy        396 --------------------------GGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 429
                                        ||||||||||||||||||||||||||||||||||
Db        421 GCTTGGACGGCTGCCGCGGTCTTGCAGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 480

Qy        430 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 489
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 540

Qy        490 TCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 549
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCCGTCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 600

Qy        550 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 609
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 660

Qy        610 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 720

Qy        670 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 780

Qy        730 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 840

Qy        790 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTC--------------------- 828
              |||||||||||||||||||||||||||||||||||||||
Db        841 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGGTACGTCGCGAGTCCTGACT 900

Qy        829 -----------------------------------------------------------G 829
                                                                         |
Db        901 GCAGTGAGCTGTTCGAGAGTCTCCCAGTTTGCTGACTGCCCTTTGTTCATGCGATTGCAG 960

Qy        830 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 1020

Qy        890 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 1080

Qy        950 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1140

Qy       1010 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCT 1200

Qy       1070 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1260

Qy       1130 GCCAGGACCGCTACCCGCTCCGCACGTCGCCT---------------------------- 1161
              ||||||||||||||||||||||||||||||||
Db       1261 GCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGGTTCGTCCCCTCTCTCTCCCCTTCC 1320

Qy       1162 -----------------------------------------CAGTTCCTCGGCCCGCTCG 1180
                                                       |||||||||||||||||||
Db       1321 CTCCGTCCGACCGGCGCGTCGAGACTTACGTTTTGCGTATCCAGTTCCTCGGCCCGCTCG 1380
```

```
Qy       1181 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1440

Qy       1241 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1500

Qy       1301 CTGTCTCGATTTCGATGGAGAAGAC----------------------------------- 1325
              |||||||||||||||||||||||||
Db       1501 CTGTCTCGATTTCGATGGAGAAGACCAGGTGCGTCTCTCGCTGCCTTCGTACTCCGATCT 1560

Qy       1326 ----------------------CAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1363
                                    ||||||||||||||||||||||||||||||||||||||
Db       1561 TGTGCTGAATGTTCTTCTCCTGCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1620

Qy       1364 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1423
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1680

Qy       1424 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1483
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1740

Qy       1484 CTTCGG------------------------------------------------------ 1489
              ||||||
Db       1741 CTTCGGAGGTGAGCCGTCGACGTTCTCCGCCGTCGCTCGTCCCCTTCAGCGCACCCAGGC 1800

Qy       1490 -------------------AGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1530
                                 |||||||||||||||||||||||||||||||||||||||||
Db       1801 TGACTTCCTTTCCCTCTGTAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1860

Qy       1531 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1920

Qy       1591 GCCGAGGCCAACGACGTCCTTTCTCTC--------------------------------- 1617
              |||||||||||||||||||||||||||
Db       1921 GCCGAGGCCAACGACGTCCTTTCTCTCGTGCGTTCGTGTCGCAATGAGTCCCGACGCAAT 1980

Qy       1618 ---------------------------CTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 1650
                                         |||||||||||||||||||||||||||||||||
Db       1981 AGCGACTGACTGCGCGATCCTGAGCAGCTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 2040

Qy       1651 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 1710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 2100

Qy       1711 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 1770
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 2160

Qy       1771 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 1830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 2220

Qy       1831 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 1890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 2280

Qy       1891 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 1950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 2340

Qy       1951 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2400

Qy       2011 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2070
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2460

Qy       2071 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2130
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2520

Qy       2131 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2580

Qy       2191 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2640

Qy       2251 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGCCTTACTCTCTCG 2310
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       2641 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGTCTTACTCTCTCG 2700

Qy       2311 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2351
              |||||||||||||||||||||||||||||||||||||||||
Db       2701 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2741
```

```
RESULT 5
AAD45824
ID   AAD45824 standard; DNA; 2163 BP.
XX
AC   AAD45824;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Yeast phenylalanine ammonia lyase DNA #2.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; ds.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2163
FT                   /*tag=  a
FT                   /product= "PAL protein"
FT                   /EC_number= "4.3.1.5"
FT                   /transl_except= (pos:13. .15, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:34. .36, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:46. .48, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:49. .51, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:55. .57, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:58. .60, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser, Leu or an in-frame stop
FT                   codon"
FT                   /transl_except= (pos:73. .75, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Pro, Ser"
FT                   /transl_except= (pos:79. .81, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro, Ser"
FT                   /transl_except= (pos:82. .84, aa:Xaa)
FT                   /note= "Xaa corresponds to Ala, Pro"
FT                   /transl_except= (pos:100. .102, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr"
FT                   /transl_except= (pos:106. .108, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Ser"
FT                   /transl_except= (pos:115. .117, aa:Xaa)
FT                   /note= "Xaa corresponds to Ala, Pro, Ser"
FT                   /transl_except= (pos:118. .120, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Gly, Trp"
FT                   /transl_except= (pos:142. .144, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Thr, Met, Glu, Ala, Val,
FT                   Gln, Pro, Leu"
FT                   /transl_except= (pos:160. .162, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
```

```
FT                   /transl_except= (pos:166. .168, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Glu, Gln"
FT                   /transl_except= (pos:193. .195, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Asp, Val"
FT                   /transl_except= (pos:196. .198, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Val, Leu"
FT                   /transl_except= (pos:226. .228, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:259. .261, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Pro, Ser"
FT                   /transl_except= (pos:277. .279, aa:Xaa)
FT                   /note= "Xaa corresponds to Asp, Ala"
FT                   /transl_except= (pos:304. .306, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Asn"
FT                   /transl_except= (pos:307. .309, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Ser"
FT                   /transl_except= (pos:325. .327, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:334. .336, aa:Xaa)
FT                   /note= "Xaa corresponds to Asp, His, Tyr"
FT                   /transl_except= (pos:340. .342, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Ser"
FT                   /transl_except= (pos:349. .351, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:442. .444, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:448. .450, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:457. .459, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Val, Phe"
FT                   /transl_except= (pos:460. .462, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Asp"
FT                   /transl_except= (pos:469. .471, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser, Gly, Arg"
FT                   /transl_except= (pos:475. .477, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:547. .549, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr"
FT                   /transl_except= (pos:667. .669, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:673. .675, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser"
FT                   /transl_except= (pos:709. .711, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser, Thr, Ile, Gly, Ala, Val,
FT                   Cys, Phe"
FT                   /transl_except= (pos:715. .717, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:721. .723, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:724. .726, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:736. .738, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Asn, Thr"
FT                   /transl_except= (pos:751. .753, aa:Xaa)
FT                   /note= "Xaa corresponds to Tyr, Ser, Phe"
FT                   /transl_except= (pos:775. .777, aa:Xaa)
FT                   /note= "Xaa corresponds to Leu, Phe"
FT                   /transl_except= (pos:913. .915, aa:Xaa)
FT                   /note= "Xaa corresponds to Leu"
FT                   /transl_except= (pos:955. .957, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1036. .1038, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Arg, Thr, Gln, Pro, Trp,
FT                   Ser or a stop codon"
FT                   /transl_except= (pos:1231. .1233, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Thr, Met"
FT                   /transl_except= (pos:1261. .1263, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1369. .1371, aa:Xaa)
FT                   /note= "Xaa corresponds to Leu"
FT                   /transl_except= (pos:1372. .1374, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1396. .1398, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1459. .1461, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:1477. .1479, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:1498. .1500, aa:Arg)
FT                   /transl_except= (pos:1552. .1554, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr"
FT                   /transl_except= (pos:1606. .1608, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:1666. .1668, aa:Xaa)
FT                   /note= "Xaa corresponds to Met, Ile, Val, Leu"
FT                   /transl_except= (pos:1669. .1671, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Val, Leu"
FT                   /transl_except= (pos:1672. .1674, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Pro"
FT                   /transl_except= (pos:1675. .1677, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1681. .1683, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Leu, Phe"
FT                   /transl_except= (pos:1684. .1686, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Glu, Gln"
FT                   /transl_except= (pos:1699. .1701, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1714. .1716, aa:Xaa)
FT                   /note= "Xaa corresponds to Asn, Tyr"
FT                   /transl_except= (pos:1723. .1725, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Ala, Val"
FT                   /transl_except= (pos:1741. .1743, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1747. .1749, aa:Xaa)
FT                   /note= "Xaa corresponds to Asn, Asp, Tyr"
FT                   /transl_except= (pos:1849. .1851, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1852. .1854, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro, Ser"
FT                   /transl_except= (pos:1927. .1929, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
FT                   stop codon"
FT                   /transl_except= (pos:1939. .1941, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Pro, Ser"
FT                   /transl_except= (pos:1948. .1950, aa:Xaa)
FT                   /note= "Xaa corresponds to Ala, Pro, Ser"
FT                   /transl_except= (pos:1978. .1980, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:2122. .2124, aa:Xaa)
FT                   /note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
FT                   Tyr, Cys"
FT                   /transl_except= (pos:2128. .2130, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser, Arg, Cys"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   P-PSDB; AAE27943.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 7; Page 53-56; 74pp; English.
XX
```

```
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.

  Query Match             70.4%;  Score 1704;  DB 6;  Length 2163;
  Best Local Similarity   84.0%;  Pred. No. 3.9e-261;
  Matches 1817;  Conservative  109;  Mismatches  236;  Indels    0;  Gaps    0;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
Db          1 ATGGCCCCCTCCBTCGACTCGATCGCGACCTCGBTCGCCAACGGCBTCDCVAACGGVTHG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 156
Db         61 CACGCCGCTCCGHCCAAGYCGSCWACGGGCGCCACGTCCACHCTCMGMCGGCCGBCCDGG 120

Qy        157 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 216
Db        121 CTCGCTCCTCCCGCCACCCAGVHGACGCAGCTCGACATCGTBGAGVAGATCCTCGCCGAC 180

Qy        217 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 276
Db        181 CCCACCGACGACGWSVTCGAACTCGACGGGTACACCCTCACCCTCGGHGACGTCGTCGGC 240

Qy        277 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 336
Db        241 GCCGCBCGCAAGGGCCGCHCBGTCCGCGTCCAGACAGMCGACGAGATCCGCGCAAAGATC 300

Qy        337 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACG 396
Db        301 GACAAVAGBGTCGAGTTCCTCCGBDCBCAGCTCBACAACAGBGTCTACGGHGTCACGACT 360

Qy        397 GGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTC 456
Db        361 GGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCVATCTCGCTCCAGAAGGCBCTC 420

Qy        457 ATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGC 516
Db        421 CTCGAGCACCAGCTCTGCGGTGTBCTCCCBACGTCGDTCGABTCCTTCVGCCTCGGHCGC 480

Qy        517 GGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAAC 576
Db        481 GGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGCGTCAAC 540

Qy        577 TCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTC 636
Db        541 TCGCTCACBCGCGGCCACTCGGCBGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTC 600

Qy        637 TTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGAC 696
Db        601 CTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCGGGCGAC 660

Qy        697 CTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCAC 756
Db        661 CTCTCCCCBCTCTCBTACATCGCCGCCGCCATCACCGGTCACCCGGACDBCAAGGTBCAC 720

Qy        757 GTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGT 816
Db        721 GTYKTSCACGAGGGCAMSGAGAAGATCATGTHCGCCCGCGAGGCGATCGCGCTCTTBGGT 780

Qy        817 CTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTC 876
Db        781 CTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACGGCCGTC 840

Qy        877 TCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAG 936
Db        841 TCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAG 900

Qy        937 GCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTC 996
Db        901 GCGCTCACGGCTCTBACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCACCCVTTC 960

Qy        997 ATCCACGACGTCTGCCGCCCGCACCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACG 1056
Db        961 CTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATCCGCACG 1020

Qy       1057 CTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGAC 1116
Db       1021 CTCCTCGAGGGCAGCHVGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAGGACGAC 1080

Qy       1117 GAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCG 1176
Db       1081 GAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTCGGCCCG 1140

Qy       1177 CTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACC 1236
Db       1141 CTCGTCAGCGACATGATTCACGCCCACGCBGTCCTCTCGCTCGAGGCCGAGTCGACGACC 1200

Qy       1237 GACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCG 1296
Db       1201 GACAACCCGCTCATCGACGTCGAGAACAAGAHGACCCACCACGGCGGCAACTTCCAGGCG 1260

Qy       1297 TCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTC 1356
Db       1261 DCCGCTGTCGCVAACACGATGGAGAAGACTCGCCTCGCVCTCGCCCTGATCGGCAAGCTC 1320

Qy       1357 AACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGC 1416
Db       1321 AACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTBCCBTCCTGC 1380

Qy       1417 CTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCT 1476
Db       1381 CTCGCTGCCGAGGACCCBTCGCTCTCCTATCACTGCAAGGGCCTCGACATTGCCGCBGCB 1440

Qy       1477 GCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCA 1536
Db       1441 GCBTACACTTCGGAGCTCGGHCACCTTGCCAACCCGGTBACGACCCACGTCCAGCCGGCH 1500

Qy       1537 GAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAG 1596
Db       1501 GAGATGGGCAACCAGGCCGTCAACTCGCTCGCBCTCATCTCGGCBCGCCGCACBGCCGAG 1560

Qy       1597 GCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTC 1656
Db       1561 GCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTBCTCCAGGCCGTC 1620

Qy       1657 GACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTC 1716
Db       1621 GACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGVTBVTCVCBDCGCTC 1680

Qy       1717 CTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAG 1776
Db       1681 HTCVAGCAGCACTTTGGCDCYGCCCTCGACGGCWACGAACTCGHGGACAAGGTCAACAAG 1740

Qy       1777 GCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGGCACGAC 1836
Db       1741 DCGCTCDACAAGCGBCTCGAGCAGACCAACTCGTACGACCTCGAGCCGCGCTGGCACGAC 1800

Qy       1837 GCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAAC 1896
Db       1801 GCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCBYCTGCCAAG 1860

Qy       1897 GTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTC 1956
Db       1861 GTCTCGCTCGCBGCCGTCAACGCCTGGAAGGTCGCCTCCGCCGAGAAGGCCATCTCGCTC 1920

Qy       1957 ACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCA 2016
Db       1921 ACGCGCBAVGTCCGCGACHCCTTCTGGBCGGCBCCGTCGTCGTCGTCGCCCGCGCTCDCG 1980

Qy       2017 TACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAG 2076
Db       1981 TACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTCCGCGAGGAGCTCGGCGTCAAG 2040

Qy       2077 GCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCG 2136
```

```
Db       2041 GCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGGTGACGATCGGCACCAACGTCTCC 2100

Qy       2137 CGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCG 2196

Db       2101 CGCATCTACGAGGCCATCAAGDVCGGCHGCATCAACCACGTCCTCGTCAAGATGCTCGCD 2160

Qy       2197 TA 2198

Db       2161 TA 2162


RESULT 6
ABA95244
ID   ABA95244 standard; DNA; 2475 BP.
XX
AC   ABA95244;
XX
DT   29-AUG-2003  (revised)
DT   07-AUG-2003  (revised)
DT   10-JUN-2002  (first entry)
XX
DE   Rhodotorula PAL consensus polynucleotide sequence.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
KW   consensus; ds.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .2475
FT                   /*tag=  a
FT                   /note= "n= A or C or G or T; "n" indicates no consensus
FT                   at that position"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   P-PSDB; ABB07694.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Claim 5; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a consensus polynucleotide sequence of the Rhodotorula PAL.
CC   (Updated on 07-AUG-2003 to correct OS field.) (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;

  Query Match             67.3%;  Score 1629;  DB 6;  Length 2475;
  Best Local Similarity   78.0%;  Pred. No. 2.9e-249;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96

Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154

Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214

Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271

Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330

Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390

Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450

Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510

Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570

Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630

Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690

Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750

Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810

Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870

Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930

Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990

Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050
```

```
Db      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy     1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110
Db     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy     1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170
Db     1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy     1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227
Db     1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy     1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287
Db     1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy     1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347
Db     1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy     1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407
Db     1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy     1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467
Db     1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy     1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527
Db     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy     1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587
Db     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy     1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647
Db     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy     1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707
Db     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy     1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC------------AACGCA 1755
Db     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy     1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815
Db     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy     1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875
Db     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy     1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933
Db     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy     1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993
Db     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy     1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053
Db     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy     2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113
Db     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy     2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173
Db     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy     2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232
Db     2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy     2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292
Db     2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy     2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352
Db     2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy     2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411
Db     2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy     2412 AAAAAAAA 2419
Db     2401 NNNANAAA 2408
```

RESULT 7
AAD45816
ID   AAD45816 standard; DNA; 2475 BP.
XX
AC   AAD45816;
XX
DT   07-AUG-2003 (revised)
DT   27-DEC-2002 (first entry)
XX
DE   Phenylalanine ammonia lyase consensus DNA.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; ds.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK, Kootstra AB;
XX
DR   WPI; 2002-690616/74.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 5; Page 45-46; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are

```
CC   also used in gene therapy. The present sequence is PAL consensus DNA.
CC   (Updated on 07-AUG-2003 to correct OS field.)
XX
SQ   Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;

  Query Match             67.3%;  Score 1629;  DB 6;  Length 2475;
  Best Local Similarity   78.0%;  Pred. No. 2.9e-249;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407
Db       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy       1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467
Db       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy       1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527
Db       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy       1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587
Db       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy       1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647
Db       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy       1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707
Db       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy       1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC-----------AACGCA 1755
Db       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740

Qy       1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815
Db       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy       1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875
Db       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy       1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933
Db       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy       1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993
Db       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy       1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053
Db       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy       2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113
```

```
Db       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy       2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173

Db       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy       2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232

Db       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy       2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292

Db       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy       2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352

Db       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy       2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411

Db       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy       2412 AAAAAAAA 2419

Db       2401 NNNANAAA 2408
```

RESULT 8
AAD33097

```
ID   AAD33097 standard; DNA; 2151 BP.
XX
AC   AAD33097;
XX
DT   01-JUL-2002  (first entry)
XX
DE   R. glutinis phenylalanine ammonia lyase (PAL) mutant DNA, RM120-2.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; mutant; variant; gene; ds.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "Rhodotorula glutinis PAL mutant"
FT                   /transl_except= (pos:592. .594, aa:Asp)
FT                   /EC_number= "4.3.1.5"
FT   mutation        replace(526, T)
FT                   /*tag=  b
FT   mutation        replace(592. .593, GG)
FT                   /*tag=  c
FT   mutation        replace(645, G)
FT                   /*tag=  d
FT   mutation        replace(792, A)
FT                   /*tag=  e
FT   mutation        replace(858, T)
FT                   /*tag=  f
FT   mutation        replace(1619, T)
FT                   /*tag=  g
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   P-PSDB; AAE20665.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Example 22; Page; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL mutant DNA. Note: This sequence is not shown in the specification,
CC   however it is constructed based on the PAL DNA shown as SEQ.ID.NO:7
CC   (AAD33075) in the sequence listing
XX
SQ   Sequence 2151 BP; 383 A; 832 C; 563 G; 373 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 1198.2;  DB 6;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 4.8e-181;
  Matches 1566;  Conservative    1;  Mismatches  524;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158

Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218

Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278

Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338

Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398

Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458

Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518

Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578

Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524
```

```
Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CCTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACCACATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 9
AAD33076
ID   AAD33076 standard; DNA; 2151 BP.
XX
AC   AAD33076;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant DNA.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; gene; mutant; variant; ds.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "Rhodotorula glutinis mutant PAL"
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
```

```
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   P-PSDB; AAE20654.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Example 10; Page 95-98; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL mutant DNA
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 8.6e-181;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664
```

```
Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 10
AAD33100

```
ID   AAD33100 standard; DNA; 2151 BP.
XX
AC   AAD33100;
XX
DT   01-JUL-2002  (first entry)
XX
DE   R. glutinis phenylalanine ammonia lyase (PAL) mutant DNA, EP18Km-6.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; mutant; variant; gene; ds.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2151
FT                   /*tag= a
FT                   /product= "Rhodotorula glutinis PAL mutant"
FT                   /EC_number= "4.3.1.5"
FT   mutation        replace(645, G)
FT                   /*tag= b
FT   mutation        replace(792, A)
FT                   /*tag= c
FT   mutation        replace(858, T)
FT                   /*tag= d
FT   mutation        replace(1619, T)
FT                   /*tag= e
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   P-PSDB; AAE20668.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Disclosure; Page; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL mutant DNA. Note: This sequence is not shown in the specification,
CC   however it is constructed based on the PAL DNA shown as SEQ.ID.NO:7
CC   (AAD33075) in the sequence listing
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 8.6e-181;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
```

```
Db    405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy    519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db    465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy    579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db    525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy    639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db    585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy    699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db    645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy    759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db    705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy    819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db    765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy    879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db    825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy    939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db    885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy    999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db    945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy   1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db   1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy   1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db   1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy   1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db   1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy   1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db   1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy   1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db   1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy   1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db   1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy   1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db   1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy   1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db   1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy   1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db   1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy   1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db   1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy   1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db   1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy   1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db   1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy   1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db   1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy   1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db   1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy   1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db   1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy   1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db   1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy   2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db   1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy   2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db   2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy   2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db   2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy   2184 CAAGATGCTCGCGTA 2198
Db   2136 CAAGATGCTCGCTTA 2150
```

RESULT 11
AAD26924

```
ID   AAD26924 standard; DNA; 2151 BP.
XX
AC   AAD26924;
XX
DT   09-APR-2002  (first entry)
XX
DE   Rhodotorula glutinis mutant phenylalanine ammonia-lyase (PAL) DNA.
XX
KW   Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
KW   para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW   Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW   aerospace application; chemical; red yeast; mutant; ds.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2151
FT                   /*tag=  a
FT                   /product= "R. glutinis mutant PAL enzyme"
FT   mutation        replace(1619, T)
FT                   /*tag=  b
XX
PN   WO200111071-A2.
XX
```

```
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US021156.
XX
PR   06-AUG-1999;   99US-0147719P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Tang X,  Vannelli TM,  Qi WW,  Sariaslani S,  Gatenby AA;
XX
DR   WPI; 2002-121549/16.
DR   P-PSDB; AAE16390.
XX
PT   Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT   comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia
PT   -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT   activity.
XX
PS   Claim 13; Page 67-68; 75pp; English.
XX
CC   The invention relates to a method for producing para-hydroxycinnamic acid
CC   (PHCA). The method comprises converting cinnamate to PHCA by converting
CC   glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC   route or by generating a new biocatalyst possessing enhanced tyrosine
CC   ammonia-lyase (TAL) activity. The method is useful for the biological
CC   production of PHCA which is useful as a monomer for the production of
CC   Liquid Crystal Polymers (LCP), where the LCP is used in electronic
CC   connectors, telecommunications and aerospace applications. LCP is also
CC   useful in medical devices, as well as chemical, and food packing
CC   applications due to its resistance to sterilising radiation. The present
CC   sequence is a DNA encoding red yeast, Rhodotorula glutinis
CC   (Rhodosporidium toruloides) mutant PAL derived by substituting Ile at
CC   position 540 with Thr. Mutant PAL enzyme posseses enhanced TAL activity
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 1196.6;  DB 6;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 8.6e-181;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
```

```
Db     1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy     1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763

Db     1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy     1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823

Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883

Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943

Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003

Db     1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063

Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123

Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183

Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198

Db     2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 12
ABS56543
ID   ABS56543 standard; DNA; 2151 BP.
XX
AC   ABS56543;
XX
DT   27-JAN-2003  (first entry)
XX
DE   R. glutinis DNA encoding PAL/TAL naturally occurring variant.
XX
KW   ds; gene; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase;
KW   PAH; para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
KW   tyrosine; cinnamic acid; PHCA; liquid crystal polymer.
XX
OS   Rhodtorula glutinis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "PAL/TAL"
XX
PN   WO200290523-A2.
XX
PD   14-NOV-2002.
XX
PF   03-MAY-2002; 2002WO-US018551.
XX
PR   04-MAY-2001; 2001US-0288701P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Qi WW,  Sariaslani FS,  Tang X;
XX
DR   WPI; 2003-058928/05.
DR   P-PSDB; ABG71255.
XX
PT   Novel recombinant host for producing para-hydroxycinnamic acid and
PT   tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT   gene encoding phenylalanine hydroxylase activity.
XX
PS   Example 4; Page 66-67; 69pp; English.
XX
CC   The invention relates to a recombinant host comprising at least one gene
CC   encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
CC   encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
CC   (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC   providing the recombinant host; and (2) growing the recombinant organism
CC   in the presence of a fermentable carbon substrate and thus producing para
CC   -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC   providing a recombinant organism comprising at least one gene encoding a
CC   phenylalanine hydroxylase activity; and (2) growing the recombinant
CC   organism in the presence of a fermentable carbon substrate to produce
CC   tyrosine. The recombinant host cell is used for producing para-
CC   hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC   production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC   monomer for the production of liquid crystal polymer. The present
CC   sequence encodes a naturally occurring variant of the R. glutinis
CC   Phenylalanine ammonia-lyase, PAL, which has TAL activity (PAL/TAL)
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 1196.6;  DB 7;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 8.6e-181;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy       99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158

Db       45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy      159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218

Db      105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy      219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278

Db      165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy      279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338

Db      225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy      339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398

Db      285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy      399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458

Db      345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy      459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518

Db      405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy      519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578

Db      465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy      579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638

Db      525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy      639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698

Db      585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy      699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
```

```
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818

Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878

Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938

Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998

Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058

Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118

Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178

Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235

Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295

Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355

Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415

Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475

Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535

Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595

Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655

Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715

Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763

Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823

Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883

Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943

Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003

Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063

Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123

Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183

Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198

Db       2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 13
ADE36055
ID   ADE36055 standard; DNA; 2151 BP.
XX
AC   ADE36055;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rhodosporidium glutinis EP18Km-6 mutant PAL encoding DNA SEQ ID NO:3.
XX
KW   recombination; nucleic acid template; recombined polypeptide;
KW   altered property; combined gene; gene; ds.
XX
OS   Rhodosporidium.
XX
PN   WO2003072743-A2.
XX
PD   04-SEP-2003.
XX
PF   26-FEB-2003; 2003WO-US005708.
XX
PR   26-FEB-2002; 2002US-0360279P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Milano J,  Tang X;
XX
DR   WPI; 2003-721766/68.
DR   P-PSDB; ADE36056.
XX
PT   Recombination of nucleic acid templates, useful e.g. for producing
PT   enzymes with altered properties, is based on cyclic extension of unpaired
PT   primers.
XX
PS   Example 2; SEQ ID NO 3; 481pp; English.
XX
CC   The present invention describes a method for the recombination of nucleic
CC   acid (NA) templates. The method comprises: (a) providing at least two
CC   double-stranded templates (T1, T2) having different 5' and 3' regions in
CC   their sense strands; (b) contacting the templates with at least one each
CC   of forward and reverse primers that anneal, respectively, only to the 3'-
CC   region of the antisense strand of T1 and to the 3'-region of the sense
CC   strand of T2; (c) extending the primers by no more than 1000 nucleotides
```

```
CC   (nt); (d) separating extended primers from their templates; (e)
CC   reannealing the extended primers, to either template; and (f) repeating
CC   cycles of steps (b)-(e) until at least one full-length extension product,
CC   i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
CC   method starting from at least one each of antisense and sense single-
CC   stranded templates; (2) generating a recombined polypeptide (I) with
CC   altered properties by expressing recombination products of the new
CC   processes and screening the polypeptides formed for properties different
CC   from those of polypeptides from either of the templates; and (3) (I)
CC   produced by method (2). The methods are useful for the recombination of
CC   nucleic acid templates of interest, and generation of a recombined
CC   polypeptide having altered properties. The method can be used to make
CC   combined genes that express proteins e.g. enzymes, cytokines, growth
CC   factors, viral proteins or microbial antigens with altered properties,
CC   e.g. stability, activity or specificity. The present sequence is used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 1196.6;  DB 9;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 8.6e-181;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895
```

```
Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003

Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063

Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123

Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183

Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198

Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 14
AAN81101
ID   AAN81101 standard; DNA; 2331 BP.
XX
AC   AAN81101;
XX
DT   27-AUG-2003 (revised)
DT   25-MAR-2003 (revised)
DT   19-NOV-1990 (first entry)
XX
DE   Sequence encoding L-phenylalanine ammonia lyase (PAL) from Rhodosporidium
DE   turoloides.
XX
KW   Cinnamic acid; expression vector; MT 10414 FerumP-8876; pYtrp6; ss.
XX
OS   Rhodosporidium toruloides.
XX
PN   EP278706-A.
XX
PD   17-AUG-1988.
XX
PF   08-FEB-1988;   88EP-00301011.
XX
PR   06-FEB-1987;   87JP-00024705.
PR   18-JUN-1987;   87JP-00152357.
XX
PA   (MITK) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N, Yoshino S, Sone S, Nakajima Y, Makiguchi N;
XX
DR   WPI; 1988-229543/33.
XX
PT   Recombinant plasmid for expression of L-phenylalanine ammonia-lyase -
PT   having combined promoter of tac promoter and Pl promoter for enhanced
PT   expression.
XX
PS   Example; Fig 2(A)-2(D); 37pp; English.
XX
CC   Plasmids were from E.coli MT 10414 (FermP-8876) carrying the plasmid
CC   pYtrp6 contg. the structural gene for PAL from Rhodosporidium toruloides.
CC   Inventors claim a recombinant plasmid contg. a DNA sequence coding for
CC   PAL with the SQ in AAP80501. Procedure for cloning the structural gene
CC   for PAL (AAN81101) is described as a reference example. The hybrid
CC   plasmid permits more efficient expression of PAL in E.coli. PAL is used
CC   for the prodn. of L-phenylalanine from cinnamic acid and ammonia (Updated
CC   on 25-MAR-2003 to correct PR field.) (Updated on 27-AUG-2003 to correct
CC   OS field.)
XX
SQ   Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;

Query Match 49.5%; Score 1196.6; DB 1; Length 2331;
Best Local Similarity 74.0%; Pred. No. 8.6e-181;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

```
Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158

Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218

Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278

Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338

Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398

Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458

Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518

Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578

Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638

Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698

Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758

Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818

Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878

Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938

Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998

Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058

Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118

Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
```

```
Db     1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy     1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db     1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy     1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db     1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy     1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db     1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy     1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db     1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy     1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db     1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy     1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db     1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy     1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db     1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy     1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db     1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy     1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db     1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy     1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db     1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy     1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db     1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198
Db     2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 15
AAN81116
ID   AAN81116 standard; DNA; 2331 BP.
XX
AC   AAN81116;
XX
DT   24-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   12-NOV-1990  (first entry)
XX
DE   L-phenylalanine ammonialyase structural gene.
XX
KW   L-phenylalanine ammonialyase; PAL; foreign gene expression;
KW   culture temperature; expression regulation; ss.
XX
OS   Rhodosporidium toruloides; (IFO 559).
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "PAL"
XX
PN   EP279665-A.
XX
PD   24-AUG-1988.
XX
PF   18-FEB-1988;   88EP-00301356.
XX
PR   19-FEB-1987;   87JP-00034397.
PR   18-JUN-1987;   87JP-00152359.
XX
PA   (MITK ) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N,  Yoshino S,  Yamamoto K,  Sone S,  Suzuki M,  Nakajima Y;
XX
DR   WPI; 1988-236895/34.
DR   P-PSDB; AAP80513.
XX
PT   Regulation of expression of foreign gene in Escherichia coli - by
PT   maintaining temp. of culture at 40 deg. C or more to suppress expression.
XX
PS   Disclosure; ); 36pp; English.
XX
CC   The cloned DNA from pSW13 isolated from the cloned cell carrying pSW13
CC   was digested with various restriction endonucleases. The sequences from
CC   the fragments were analysed by the method of Maxam-Gilbert and also by
CC   biochemical means using the method of Maat. The sequences obtained were
CC   edited using the GENAS program produced by Mitsui Information Development
CC   Co. The sequence so determined is represented here. The regulation of
CC   expressiom of this foreign gene in E.coli carrying a recombinant plasmid
CC   comprises maintaining the culture at at least 40 deg. C. See also EP-
CC   279664. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
CC   2003 to correct PI field.) (Updated on 24-OCT-2003 to standardise OS
CC   field)
XX
SQ   Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;

  Query Match             49.5%;  Score 1196.6;  DB 1;  Length 2331;
  Best Local Similarity   74.0%;  Pred. No. 8.6e-181;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
```

```
Db   165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy   279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db   225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy   339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db   285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy   399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db   345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy   459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db   405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy   519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db   465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy   579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db   525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy   639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db   585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy   699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db   645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy   759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db   705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy   819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db   765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy   879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db   825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy   939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db   885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy   999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db   945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy  1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db  1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy  1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db  1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy  1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db  1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy  1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db  1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy  1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db  1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy  1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db  1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy  1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db  1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy  1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db  1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy  1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db  1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy  1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db  1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy  1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db  1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy  1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db  1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy  1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db  1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy  1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db  1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy  1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db  1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy  1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db  1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy  2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db  1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy  2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db  2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy  2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db  2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy  2184 CAAGATGCTCGCGTA 2198
Db  2136 CAAGATGCTCGCTTA 2150
```

Search completed: September 11, 2004, 14:50:22
Job time : 1351.06 secs

```
OM nucleic - nucleic search, using sw model

Run on:         September 11, 2004, 13:38:08 ; Search time 181.4 Seconds
                                          (without alignments)
                                          7400.353 Million cell updates/sec

Title:          US-09-939-408A-12
Perfect score:  2419
Sequence:       1 ctcctgcctcactctaaccc..........gagttcaaaaaaaaaaaaaa 2419

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2418.6	100.0	2419	4	US-09-624-693A-12	Sequence 12, Appl
2	1629	67.3	2475	4	US-09-624-693A-20	Sequence 20, Appl
3	1196.6	49.5	2151	4	US-09-627-216A-9	Sequence 9, Appli
4	1196.6	49.5	2151	4	US-09-765-873A-9	Sequence 9, Appli
5	1196.6	49.5	2439	4	US-09-624-693A-18	Sequence 18, Appl
6	1195	49.4	2151	4	US-09-627-216A-7	Sequence 7, Appli
7	1195	49.4	2151	4	US-09-765-873A-7	Sequence 7, Appli
8	1188.8	49.1	2061	4	US-09-765-873A-31	Sequence 31, Appl
9	1136.6	47.0	2787	4	US-09-624-693A-16	Sequence 16, Appl
10	302.8	12.5	2311	4	US-09-624-693A-14	Sequence 14, Appl
11	207.2	8.6	1455	4	US-09-615-192A-245	Sequence 245, App
12	168.6	7.0	1520	4	US-09-615-192A-247	Sequence 247, App
13	126.8	5.2	1548	4	US-09-252-991A-16475	Sequence 16475, A
c 14	126.8	5.2	2295	4	US-09-252-991A-16126	Sequence 16126, A
15	108.4	4.5	1527	4	US-09-489-039A-6404	Sequence 6404, Ap
16	101.6	4.2	577	4	US-09-615-192A-97	Sequence 97, Appl
17	101.6	4.2	577	4	US-09-169-789-97	Sequence 97, Appl
18	101.6	4.2	684	2	US-08-975-316-45	Sequence 45, Appl
19	101.6	4.2	684	4	US-09-615-192A-45	Sequence 45, Appl
20	101.6	4.2	684	4	US-09-169-789-45	Sequence 45, Appl
21	95.8	4.0	681	4	US-09-615-192A-244	Sequence 244, App
22	87.6	3.6	624	2	US-08-713-000-9	Sequence 9, Appli
23	87.6	3.6	624	2	US-08-975-316-9	Sequence 9, Appli
24	87.6	3.6	624	3	US-09-211-710-9	Sequence 9, Appli
25	87.6	3.6	624	4	US-09-615-192A-9	Sequence 9, Appli
26	87.6	3.6	624	4	US-09-169-789-9	Sequence 9, Appli
27	79.4	3.3	1926	4	US-09-249-585A-4	Sequence 4, Appli
28	79.4	3.3	1931	2	US-09-130-114-2	Sequence 2, Appli
c 29	79.2	3.3	1557	4	US-09-252-991A-16124	Sequence 16124, A
30	78.8	3.3	1563	4	US-09-252-991A-16478	Sequence 16478, A
31	75.2	3.1	278	2	US-08-713-000-10	Sequence 10, Appl
32	75.2	3.1	278	2	US-08-975-316-10	Sequence 10, Appl
33	75.2	3.1	278	3	US-09-211-710-10	Sequence 10, Appl
34	75.2	3.1	278	4	US-09-615-192A-10	Sequence 10, Appl
35	75.2	3.1	278	4	US-09-169-789-10	Sequence 10, Appl
36	73.6	3.0	315	4	US-09-252-991A-16252	Sequence 16252, A
37	72	3.0	2064	1	US-08-343-428-1	Sequence 1, Appli
38	71	2.9	13842	3	US-09-105-537-30	Sequence 30, Appl
39	71	2.9	36778	3	US-09-105-537-5	Sequence 5, Appli
40	70.4	2.9	38506	3	US-09-320-878-19	Sequence 19, Appl
41	70.4	2.9	38506	4	US-09-141-908-1	Sequence 1, Appli
42	70.4	2.9	38506	4	US-09-657-440-19	Sequence 19, Appl
43	67.8	2.8	492	4	US-09-615-192A-98	Sequence 98, Appl
44	67.8	2.8	492	4	US-09-169-789-98	Sequence 98, Appl
45	64.2	2.7	294	4	US-09-615-192A-246	Sequence 246, App

ALIGNMENTS

```
RESULT 1
US-09-624-693A-12
; Sequence 12, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSC
;  CURRENT APPLICATION NUMBER: US/09/624,693A
;  CURRENT FILING DATE:  2000-07-24
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 2419
;   TYPE: DNA
;   ORGANISM: Rhodotorula graminis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (37)..(2196)
US-09-624-693A-12

  Query Match             100.0%;  Score 2418.6;  DB 4;  Length 2419;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2418;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60

Qy         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120

Qy        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180

Qy        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240

Qy        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300

Qy        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360

Qy        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420

Qy        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480

Qy        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGCCGACGTCCGTCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540

Qy        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600

Qy        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660

Qy        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720

Qy        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780

Qy        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840

Qy        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900

Qy        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960

Qy        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020

Qy       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080

Qy       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140

Qy       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200

Qy       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260

Qy       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320

Qy       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380

Qy       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440

Qy       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500

Qy       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560

Qy       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620

Qy       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680

Qy       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740

Qy       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800

Qy       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860

Qy       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920

Qy       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980

Qy       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040

Qy       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100

Qy       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160

Qy       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220

Qy       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280

Qy       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340

Qy       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400

Qy       2401 AGTTCAAAAAAAAAAAAAA 2419
              |||||||||||||||||||
Db       2401 AGTTCAAAAAAAAAAAAAA 2419
```

RESULT 2
US-09-624-693A-20
; Sequence 20, Application US/09624693A

```
; Patent No. 6355468
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSC
;  CURRENT APPLICATION NUMBER: US/09/624,693A
;  CURRENT FILING DATE:  2000-07-24
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 2475
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_difference
;   LOCATION: (13, 34, 46, 49, 51, 57, 59, 68, 69, 73, 75 - 77, 79, 82, 84,
;   OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus
;   OTHER INFORMATION: Sequence of SEQ ID NOs: 12, 16, and 18
; Patent No. 6355468
US-09-624-693A-20

  Query Match             67.3%;  Score 1629;  DB 4;  Length 2475;
  Best Local Similarity   78.0%;  Pred. No. 7.3e-307;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407
Db       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy       1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467
Db       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy       1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527
Db       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy       1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587
Db       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy       1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647
Db       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy       1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707
Db       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy       1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC------------AACGCA 1755
Db       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy       1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815
```

```
Db      1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy      1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875
Db      1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy      1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933
Db      1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy      1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993
Db      1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy      1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053
Db      1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy      2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113
Db      2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy      2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173
Db      2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy      2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232
Db      2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy      2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292
Db      2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy      2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352
Db      2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy      2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411
Db      2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy      2412 AAAAAAAA 2419
Db      2401 NNNANAAA 2408
```

RESULT 3
US-09-627-216A-9
; Sequence 9, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mutant from
; OTHER INFORMATION: Rhodotorula glutinis

US-09-627-216A-9

Query Match 49.5%; Score 1196.6; DB 4; Length 2151;
Best Local Similarity 74.0%; Pred. No. 3.8e-223;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

```
Qy        99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db        45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy       159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db       105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy       219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db       165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy       279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db       225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy       339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db       285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy       399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db       345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy       459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db       405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy       519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db       465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy       579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db       525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy       639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db       585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy       699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db       645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy       759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db       705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy       819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db       765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy       879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db       825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy       939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db       885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy       999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db       945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy      1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db      1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064
```

```
Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 4
US-09-765-873A-9
; Sequence 9, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-9

```
  Query Match             49.5%;  Score 1196.6;  DB 4;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 3.8e-223;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644
```

```
Qy    699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db    645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy    759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db    705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy    819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db    765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy    879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db    825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy    939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db    885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy    999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db    945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy   1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db   1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy   1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db   1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy   1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db   1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy   1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db   1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy   1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db   1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy   1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db   1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy   1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db   1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy   1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db   1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy   1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db   1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy   1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db   1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy   1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db   1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy   1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db   1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy   1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db   1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy   1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db   1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy   1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db   1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy   1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db   1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy   2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db   1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy   2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db   2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy   2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db   2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy   2184 CAAGATGCTCGCGTA 2198
Db   2136 CAAGATGCTCGCTTA 2150
```

RESULT 5
US-09-624-693A-18
; Sequence 18, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Rhodotorula toruloides
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2148)
US-09-624-693A-18

```
  Query Match             49.5%;  Score 1196.6;  DB 4;  Length 2439;
  Best Local Similarity   74.0%;  Pred. No. 3.9e-223;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy     99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db     45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy    159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db    105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy    219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db    165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224
```

```
Qy    279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db    225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy    339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db    285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy    399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db    345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy    459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db    405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy    519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db    465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy    579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db    525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy    639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGCGACCT 698
Db    585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy    699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db    645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy    759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db    705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy    819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db    765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy    879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db    825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy    939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db    885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy    999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db    945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy   1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db   1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy   1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db   1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy   1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db   1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy   1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db   1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy   1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db   1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy   1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db   1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy   1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db   1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy   1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db   1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy   1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db   1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy   1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db   1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy   1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db   1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy   1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db   1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy   1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db   1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy   1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db   1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy   1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db   1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy   1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db   1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy   2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db   1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy   2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db   2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy   2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db   2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy   2184 CAAGATGCTCGCGTA 2198
Db   2136 CAAGATGCTCGCTTA 2150
```

RESULT 6
US-09-627-216A-7
; Sequence 7, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A

```
;  CURRENT FILING DATE:  2000-07-27
;  PRIOR APPLICATION NUMBER: 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis
US-09-627-216A-7

  Query Match             49.4%;  Score 1195;  DB 4;  Length 2151;
  Best Local Similarity   73.9%;  Pred. No. 7.8e-223;
  Matches 1564;  Conservative    1;  Mismatches  526;  Indels   24;  Gaps    3;

Qy          99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db          45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy         159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db         105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy         219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db         165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy         279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db         225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy         339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db         285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy         399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db         345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy         459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db         405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy         519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db         465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy         579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db         525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy         639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db         585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy         699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db         645 GTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy         759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db         705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy         819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db         765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy         879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db         825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTCTCGCAGTC 884

Qy         939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db         885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy         999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db         945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy        1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db        1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy        1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db        1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy        1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db        1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy        1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db        1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy        1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db        1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy        1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db        1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy        1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db        1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy        1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db        1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy        1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db        1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy        1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db        1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy        1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db        1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy        1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db        1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy        1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db        1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy        1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db        1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy        1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db        1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy        1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db        1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy        2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db        1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015
```

```
Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123

Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183

Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198

Db       2136 CAAGATGCTCGCTTA 2150


RESULT 7
US-09-765-873A-7
; Sequence 7, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2151)
US-09-765-873A-7

  Query Match             49.4%;  Score 1195;  DB 4;  Length 2151;
  Best Local Similarity   73.9%;  Pred. No. 7.8e-223;
  Matches 1564;  Conservative    1;  Mismatches  526;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158

Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218

Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278

Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338

Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398

Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458

Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518

Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578

Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638

Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698

Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758

Db        645 GTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818

Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878

Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938

Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998

Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058

Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118

Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178

Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235

Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295

Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355

Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415

Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475

Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535

Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595

Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655

Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604
```

```
Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 8

US-09-765-873A-31

```
; Sequence 31, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
;   LENGTH: 2061
;   TYPE: DNA
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: CDS]
;   LOCATION: (1)..(2061)
US-09-765-873A-31

  Query Match             49.1%;  Score 1188.8;  DB 4;  Length 2061;
  Best Local Similarity   75.2%;  Pred. No. 1.2e-221;
  Matches 1529;  Conservative    1;  Mismatches  478;  Indels   24;  Gaps    3;

Qy        182 CGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTCA 241
Db         38 CGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCCGACCGACTCGACGCTCGAACTCG 97

Qy        242 GCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTCC 301
Db         98 ACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGCCGCGAGGAAGGGCAGGCCTGTCC 157

Qy        302 GCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAGG 361
Db        158 GCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGACAAATCGGTCGAGTTCTTGCGCT 217

Qy        362 CCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACGA 421
Db        218 CGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGGATTTGGCGGATCCGCAGACACCC 277

Qy        422 GGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGA 481
Db        278 GCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCTCGAGCACCAGCTCTGCGGTGTTC 337

Qy        482 CGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCG 541
Db        338 TCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGGTCTCGAGAACTCGCTTCCCCTCG 397

Qy        542 AGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCG 601
Db        398 AGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAGCTTGACCCGCGGCCACTCGGCTG 457

Qy        602 TCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCG 661
Db        458 TCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCTCAACCACGGCATCACCCCCATCG 517

Qy        662 TCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCG 721
Db        518 TCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCTGTCTCCTCTCTCCTACATTGCAG 577

Qy        722 GCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGA 781
Db        578 CGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGTCGTCCACGAGGGCAAGGAGAAGA 637

Qy        782 TCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCGA 841
Db        638 TCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCTCGAGCCCGTCGTCCTCGGCCCGA 697

Qy        842 AGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGTC 901
Db        698 AGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTCAGCATCGATGGCCACCCTCGCTC 757

Qy        902 TGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAGG 961
Db        758 TGCACGACGCTCACATGCTCTCGCTCCTCTCGCAGTCGCTCACGGCCATGACGGTCGAAG 817

Qy        962 CCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCACC 1021
Db        818 CGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCTTCACGACGTCACGCGCCCTCACC 877

Qy       1022 CCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCCG 1081
Db        878 CGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCTCCTCGAGGGAAGCCGCTTTGCTG 937

Qy       1082 TTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGCT 1141
Db        938 TCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTCCGCCAGGACCGCT 997

Qy       1142 ACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCCT 1201
Db        998 ACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCTCGTCAGCGACCTCATTCACGCCC 1057

Qy       1202 ACTCGACTCTCTCGCTCGAG---AACAACACGACGACCGACAACCCGCTCCTCGACGTCG 1258
Db       1058 ACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGACCGACAACCCTCTCATCGACGTCG 1117

Qy       1259 AGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGG 1318
Db       1118 AGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGCTGCCGCTGTGGCCAACACCATGG 1177
```

```
Qy       1319 AGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGT 1378
Db       1178 AGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCTCAACTTCACGCAGCTCACCGAGA 1237

Qy       1379 TGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGC 1438
Db       1238 TGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTGCCTCGCGGCCGAAGACCCCTCGC 1297

Qy       1439 TCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCC 1498
Db       1298 TCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGCGGCGTACACCTCGGAGTTGGGAC 1357

Qy       1499 ACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCA 1558
Db       1358 ACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGCTGAGATGGCGAACCAGGCGGTCA 1417

Qy       1559 ACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCC 1618
Db       1418 ACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGAGTCCAACGACGTCCTTTCTCTCC 1477

Qy       1619 TTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCG 1678
Db       1478 TCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCATCGACTTGCGCGCGATCGAGTTCG 1537

Qy       1679 ACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTG 1738
Db       1538 AGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCTCATCGACCAGCACTTTGGCTCCG 1597

Qy       1739 GCCTCGACG------------TCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACA 1786
Db       1598 CCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGAGAAGGTGAACAAGACGCTCGCCA 1657

Qy       1787 AGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGT 1846
Db       1658 AGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCCGCGCTGGCACGACGCCTTCTCCT 1717

Qy       1847 ACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTA 1906
Db       1718 TCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGACGTC---------GCTCTCGCTCG 1768

Qy       1907 CTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGG 1966
Db       1769 CCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTCGGCCATCTCGCTCACCCGCCAAG 1828

Qy       1967 TGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGC 2026
Db       1829 TCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTCGCCCGCGCTCTCGTACCTCTCGC 1888

Qy       2027 CGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCG 2086
Db       1889 CGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCG 1948

Qy       2087 GCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACG 2146
Db       1949 GAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGGCTCGAACGTCTCCAAGATCTACG 2008

Qy       2147 AGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTA 2198
Db       2009 AGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCTCAAGATGCTCGCTTA 2060
```

RESULT 9
US-09-624-693A-16
; Sequence 16, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 2787
; TYPE: DNA
; ORGANISM: Rhodotorula mucilaginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (646)..(2784)
US-09-624-693A-16

```
  Query Match             47.0%;  Score 1136.6;  DB 4;  Length 2787;
  Best Local Similarity   71.6%;  Pred. No. 1.6e-211;
  Matches 1557;  Conservative    1;  Mismatches  590;  Indels   27;  Gaps    4;

Qy         27 AGCACTCGCGATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTAC 86
Db        636 AGCTCTCAAGATGGCCCCCTCCGTCGACTCGATCGCGACTTCGGTTGCCAACTCCCTCTC 695

Qy         87 CAACGGCTCGCACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCG 146
Db        696 GAACGGGTTGCACGCCGCCGCCGCCGCCAACGGTGGCGACGTCCACAAGAAGACGGCCGG 755

Qy        147 CACGCCCGGCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCT 206
Db        756 TGCTGGCTCCCTCCTCCCGACCACCGAGACGACCCAGCTCGACATCGTTGAGCGCATCTT 815

Qy        207 CCTCAGCGACCCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGA 266
Db        816 GGCCGACGCCGGCGCGACGGACCAGATCAAACTCGATGGGTACACCCTCACGCTCGGCGA 875

Qy        267 CGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCG 326
Db        876 CGTCGTCGGCGCTGCTCGCCGTGGCCGCTCCGTCAAGGTCGCAGACAGCCCGCACATCCG 935

Qy        327 CGCACGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGG 386
Db        936 CGAGAAGATCGATGCCAGTGTCGAGTTCCTCCGTACTCAGCTCGACAACAGTGTCTACGG 995

Qy        387 AGTCACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCA 446
Db        996 TGTCACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCGATCTCGCTCCA 1055

Qy        447 GAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAG 506
Db       1056 AAAGGCCCTGCTCGAGCACCAGCTCTGCGGTGTCCTCCCCACCTCGATGGATGGCTTTGC 1115

Qy        507 CGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCAT 566
Db       1116 GCTCGGTCGCGGCCTCGAGAACTCGCTTCCGCTCGAAGTCGTCCGAGGCGCGATGACCAT 1175

Qy        567 CCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCT 626
Db       1176 CCGTGTCAACTCGCTCACTCGCGGTCACTCGGCGGTCCGCATCGTCGTCCTCGAAGCCCT 1235

Qy        627 CACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGC 686
Db       1236 CACCAACTTCCTCAACCACGGCATCACCCCGATCGTCCCGCTTCGAGGCACCATCTCGGC 1295

Qy        687 GTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGT 746
Db       1296 GTCGGGCGACCTTTCCCCCCTCTCTTACATCGCCGCCTCGATCACCGGCCACCCGGACTC 1355

Qy        747 CAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTC 806
Db       1356 GAAGGTCCACGT------------CGACGGCAAGATCATGTCCGCCCAGGAGGCGATCGC 1403

Qy        807 GCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGG 866
Db       1404 GCTCAAGGGTCTTCAGCCCGTCGTCCTCGGTCCGAAGGAGGGTCTCGGTCTCGTCAACGG 1463
```

```
Qy        867 AACGGCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCT 926
Db       1464 CACGGCCGTCTCCGCCTCGATGGCGACGCTGGCCCTCACCGACGCACACGTCCTCTCGCT 1523
Qy        927 CCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTT 986
Db       1524 CCTCGCACAGGCGCTCACTGCTCTTACTGTCGAGGCCATGGTCGGACACGCCGGCTCGTT 1583
Qy        987 CGCGCCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAA 1046
Db       1584 CCACCCATTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTGGCGCGCAA 1643
Qy       1047 CATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGT 1106
Db       1644 CATCCGGACTCTTCTCGAGGGCAGCAAGTACGCCGTCCACCACGAGACTGAAGTCAAGGT 1703
Qy       1107 CAAGGACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTT 1166
Db       1704 CAAGGACGACGAGGGCATCCTCAGGCAGGACCGGTACCCGCTCCGCTGCTCGCCGCAGTG 1763
Qy       1167 CCTCGGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAG---AA 1223
Db       1764 GCTCGGTCCCCTTGTCAGCGACATGATTCACGCTCACGCTGTCCTCTCGCTCGAGGCTGG 1823
Qy       1224 CAACACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGG 1283
Db       1824 TCAGTCGACCACCGACAACCCGCTGATCGACCTCGAGAACAAGATGACCCACCATGGCGG 1883
Qy       1284 CAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCT 1343
Db       1884 AGCCTTCATGGCGAGCAGCGTCGGAAACACGATGGAGAAGACTCGCCTCGCCGTCGCGCT 1943
Qy       1344 CATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGG 1403
Db       1944 GATGGGCAAGGTCAGCTTTACTCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGGGC 2003
Qy       1404 CCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGA 1463
Db       2004 CCTTCCGTCCTGCCTCGCTGCCGAGGACCCTTCCCTCTCTTATCACTGCAAGGGTCTCGA 2063
Qy       1464 CATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTT 1523
Db       2064 CATTGCTGCGGCCGCCTACACTTCCGAGCTCGGTCACCTTGCCAACCCGGTTTCGACCCA 2123
Qy       1524 CGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCG 1583
Db       2124 CGTCCAGCCGGCCGAGATGGGCAACCAGGCCATCAACTCGCTCGCCCTCATCTCGGCCCG 2183
Qy       1584 CCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCAC 1643
Db       2184 CCGCACCGCCGAGGCGAACGACGTTCTCTCCCTCCTCCTCGCCACCCACCTCTACTGCGT 2243
Qy       1644 GCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCT 1703
Db       2244 CCTCCAGGCCGTCGACCTCCGCGCGATGGAGTTTGAGCACACCAAGGCGTTCGAGCCGAT 2303
Qy       1704 TCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCT 1763
Db       2304 GGTCACTGAGCTGTTGAAGCAGCACTTTGGC---GCGCTCGCGACGGCCGAAGTCGAGGA 2360
Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       2361 CAAGGTCCGCAAGTCGATCTACAAGCGGTTGCAGCAGAACAACTCGTACGACCTCGAGCA 2420
Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       2421 GCGGTGGCACGACACGTTCTCGGTCGCGACCGGTGCCGTCGTCGAGGCGCTCGCCGGCCA 2480
Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       2481 G---------GAGGTCTCGCTCGCGAGCCTCAACGCCTGGAAGGTCGCCTGCGCCGAGAA 2531
Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       2532 GGCTATCGCGCTCACGCGCTCCGTCCGCGACTCGTTCTGGGCGGCTCCGTCGTCGTCGTC 2591
Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       2592 GCCCGCGCTCAAGTACCTCTCCCCGCGGACGCGCGTCCTGTATTCGTTCGTCCGGGAGGA 2651
Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2652 GGTCGGCGTCAAGGCCCGCCGCGGCGATGTCTACCTCGGCAAGCAGGAGGTCACGATCGG 2711
Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2712 CACCAACGTCAGCCGCATCTACGAGGCGATCAAGAGCGGTTGCATCGCCCCCGTCCTCGT 2771
Qy       2184 CAAGATGCTCGCGTA 2198
Db       2772 CAAGATGATGGCATA 2786
```

RESULT 10
US-09-624-693A-14
; Sequence 14, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 2311
;   TYPE: DNA
;   ORGANISM: Amanita muscaria
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (18)..(2237)
US-09-624-693A-14

```
  Query Match             12.5%;  Score 302.8;  DB 4;  Length 2311;
  Best Local Similarity   54.2%;  Pred. No. 4.5e-50;
  Matches  719;  Conservative    1;  Mismatches  573;  Indels   33;  Gaps    4;

Qy        379 GTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTC 438
Db        372 GTATACGGACTGTCAACTGGTTTCGGTGGCAGTGCTGATACACGGACGGACAAACCGATG 431
Qy        439 AGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCG 498
Db        432 TTGTTGGGGTTTGCCCTTTTGCAACACCAACATGTAGGGATACTGCCCACCTCGACTGAG 491
Qy        499 TCCTTCAGCGTCGGACGCGGCCTCG---------AGAACACGCTTCCGCTCGAGGTCGTC 549
Db        492 CCTTTGGACGTCCTACCTCTCCAAGATGCAAATAACACAAGCATGCCAGAGGCGTGGATT 551
Qy        550 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 609
Db        552 CGCGGGGCCATTTTGATCCGTATGAATTCGCTAATTCGTGGCCACTCTGGAATCAGATGG 611
Qy        610 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 669
Db        612 GAGTTGATCGAAAAGATGAGAGAACTACTCGCGGCCAATGTGATACCTGTCGTTCCCCTG 671
Qy        670 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 729
Db        672 AGAGGCAGCATCTCCTCATCCGGAGATCTGTCTCCCCTATCCTATATCGCAGGCACGATT 731
Qy        730 ACCGGTCACCCCGACGTCAAGGTTCA------CGTTTGCACGAGGGAACCGAGAAGATC 783
```

```
Db    732 ATTGGCAACCCATCAATCAAGGTATATCACGGTCCATCAAAGTCCGGAATTCGCCAAATT 791

Qy    784 ATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAG 843
Db    792 GGATCCTCGAAGGATGTCTTGGCTCTGCATAATATCGAACCTTTCCCACTGGAATCGAAA 851

Qy    844 GAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGTCTG 903
Db    852 GAACCTCTTGGTATTTGAATGGGACCGCATTCTCGGCATCTGTGGCAGCTTTAGCCCTA 911

Qy    904 CACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAGGCC 963
Db    912 AACGAAGCTATCCATCTTGTCTTGTTGGCTCAAGTGTGCACGGCTATGGGGACCGAGGCA 971

Qy    964 ATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCACCCC 1023
Db    972 TTGATAGGCACTCGCGCTTCTCATGCACCGTTCATTCATGCCACCGCACGACCACATCCC 1031

Qy   1024 GGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTT 1083
Db   1032 GGTCAAGTAGAATGTGCTGAGAACATTTGGAATTTGCTCGATGGGAGTAAATTGGCTCAG 1091

Qy   1084 GAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGCTAC 1143
Db   1092 TTAGAAGAGCACGAAGTTCGCCTAGAAGACGATAAATACACCCTTCGGCAGGACCGTTAT 1151

Qy   1144 CCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCCTAC 1203
Db   1152 CCACTCCGAACTTCGCCTCAATTCCTTGGGCCTCAGATTGAAGACATAATCTCCGCTTTC 1211

Qy   1204 TCGACTCTCTCGCTCGAG------AACAACACGACGACCGACAACCCGCTCCTCGACGTC 1257
Db   1212 CAGACTGTAACGCAGGAGTGTAATTACTTACCAGCTACTGACAATCCACTGATTGATGGG 1271

Qy   1258 GAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATG 1317
Db   1272 GAGACTGGCGAATCTCACCACGGTGGCAATTTCCAAGCGATGGCTGTAACTAATGCAATG 1331

Qy   1318 GAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAG 1377
Db   1332 GAGAAGACGCGACTTGCTTTACATCACGTTGGCAAATTACTATTTTCCCAGAGCACTGAA 1391

Qy   1378 TTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCG 1437
Db   1392 TTAGTCAATCCTGCGATGAACCGCGGTCTGCCGCCTTCAGTAGCTGCCACAGATCCATCT 1451

Qy   1438 CTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGC 1497
Db   1452 CTCAACTACCACGCCAAAGGACTAGACATAGCAACTGCGGCCTACGTAGCCGA------- 1504

Qy   1498 CACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTC 1557
Db   1505 -----AGCGACTCCTGGCCCCACTCACATTCAGTCGGCAGAAATGCACAACCAAGCTGTT 1559

Qy   1558 AACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTC 1617
Db   1560 AACTCCCTGGCGTTGATTTCTGCTCGGGCTACCATCACATCGTTGGAAGTGCTAACATCT 1619

Qy   1618 CTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTC 1677
Db   1620 CTGATCGCGTCTTACTTGTATATTCTATGCCAAGCTCTCGACCTCCGTGCCCTTCAGCGC 1679

Qy   1678 GACTTC 1683
Db   1680 GAGTTC 1685
```

RESULT 11
US-09-615-192A-245
; Sequence 245, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 245
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-615-192A-245

```
Query Match             8.6%;  Score 207.2;  DB 4;  Length 1455;
Best Local Similarity   52.4%;  Pred. No. 1.4e-31;
Matches  623;  Conservative    0;  Mismatches  528;  Indels   39;  Gaps    6;

Qy    184 CAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTCAGC 243
Db    299 CACCTCGACGAGGTGAAGCGGATGGTCGAGGAGTACCGGAGGCCGGCGGTGCGCCTCGGC 358

Qy    244 GGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGC 303
Db    359 GGGGAGTCCCTCACGATAGCCCAGGTGGCGGCGGTGGCGAGTCAGGAG---GGGGTAGGG 415

Qy    304 GTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAGGCC 363
Db    416 GTCGAGCTCTCGGAGGCGGCCCGTCCCAGGGTCAAGGCCAGCAGCGACTGGGTCATGGAG 475

Qy    364 CAGCTTCAGAA------CTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGAC 417
Db    476 AGCATGAACAAGGGAACTGACAGCTACGGGGTCACCACGGGTTCGGCGCCACTTCTCAC 535

Qy    418 ACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGC 477
Db    536 CGGAGGACGAAGCAAGGCGGTGCTTTGCAGAAGGAACTTATAAGGTTCTTGAATGCCGG- 594

Qy    478 GTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCG 537
Db    595 -------------------GATCTTCGGCAACGGCACGGAGTCGTGCCACACCCTGCCT 634

Qy    538 CTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCG 597
Db    635 CAATCCTCCACCCGAGCCGCCATGCTCGTCCGGGTCAACACCCTCCTCCAGGGCTACTCC 694

Qy    598 GCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCC 657
Db    695 GGCATCCGTTTTGAGATCCTCGAGGCCATCACCAAGTTCCTCAACCACAACATCACCCCG 754

Qy    658 ATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATC 717
Db    755 TGCCTGCCCCTCAGGGGCACCATCACTGCCTCAGGCGACTTGGTCCCCCTCTCCTACATT 814

Qy    718 GCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTGCACGAGGGAACCGAG 777
Db    815 GCCGGGCTCCTGACGGGCCGGCCCAACTCCAAGG---CCGTCGGGCCTGATGGGAAGTCC 871

Qy    778 AAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGC 837
Db    872 CTGGACGCTGTCGAGGCCTTCCGGCTCGCCGGGATTGACACGGGCTTCTTCGAGCTGCAG 931

Qy    838 CCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTC 897
Db    932 CCAAAGGAAGGGTTGGCGCTCGTGAACGGCACGGCAGTCGGGTCTGGCCTGGCTTCCATC 991
```

```
Qy        898 AGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTG 957
Db        992 GTCCTCTTCGAGGCCAACATACTCGCGGTCCTGTCCGAGGTCCTGTCAGCGATCTTCGCA 1051
Qy        958 GAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCG 1017
Db       1052 GAGGTGATGCAGGGGAAGCCGGAGTTCACAGACCACTTGACGCATAAATTGAAGCAC--- 1108
Qy       1018 CACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTT 1077
Db       1109 CATCCCGGGCAGATTGAGTCTGCGGCTATAATGGA---GCACATTTTGGATGGAAGCGCT 1165
Qy       1078 GCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGAC 1137
Db       1166 TACGTGAAGGCTGCTAAAAAGTTGCACGAGATGGATCCGCTCCAGAAGCCAAAGCAGGAC 1225
Qy       1138 CGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCAC 1197
Db       1226 AGGTACGCTCTCAGGACTTCTCCCCAGTGGCTAGGGCCCCAGATTGAGGTGATCCGAGCG 1285
Qy       1198 GCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTC 1257
Db       1286 GCAACCAAGATGATTGAGAGGGAAATCAATTCGGTCAATGACAACCCGCTGATCGATGTC 1345
Qy       1258 GAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATG 1317
Db       1346 GCGAGGAACAAGGCCCTGCACGGTGGGAACTTCCAGGGGACCCCGATTGGTGTCTCCATG 1405
Qy       1318 GAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCA 1367
Db       1406 GACAACACTCGCCTGGCGGTTGCGTCCATAGGGAAGCTCATGTTCGCGCA 1455
```

RESULT 12
US-09-615-192A-247
; Sequence 247, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 247
; LENGTH: 1520
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-615-192A-247

```
  Query Match             7.0%;  Score 168.6;  DB 4;  Length 1520;
  Best Local Similarity  51.4%;  Pred. No. 4.1e-24;
  Matches  469;  Conservative    0;  Mismatches  434;  Indels   10;  Gaps    3;

Qy        838 CCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTC 897
Db        141 CCTAAAGAAGGTCTGGCCATTGTCAATGGCACTTCAGTGGAGCAGCACTGGCTTCCATT 200
Qy        898 AGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTG 957
Db        201 GTGTGTTTCGATGCCAATGTTCTTGCTCTGCTCTCTGAAGTAATCTCTGCCATGTTCTGC 260
Qy        958 GAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCG 1017
Db        261 GAGGTTATGAATGGGAAGCCTGAGTTTACAGATCCATTAACTCACAAGCTGAAGCAC--- 317
Qy       1018 CACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTT 1077
Db        318 CATCCTGGCCAAATGGAAGCTGCAGCGATCATGGAGTATGTCTT----GGACGGGAGTCT 373
Qy       1078 GCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGAC 1137
Db        374 TATATGAAACACGCTGCTAAGCTCCATGAGATGAATCCTCTGCAGAAGCCAAAGCAGGAT 433
Qy       1138 CGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCAC 1197
Db        434 CGCTATGCGCTTCGCACTTCGCCTCAGTGGCTCGGCCCTCAGGTGGAGATTATCAGATCT 493
Qy       1198 GCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTC 1257
Db        494 GCAACTCACATGATTGAGCGGGAAATCAATTCTGTGAATGACAATCCAGTAATTGATGTT 553
Qy       1258 GAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATG 1317
Db        554 GCCAGAGACAAAGCTCTACATGGAGGGAATTTCCAGGGCACACCTATTGGTGTTTCCATG 613
Qy       1318 GAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAG 1377
Db        614 GATAATCTTCGTCTGTCAATTTCAGCAATTGGGAAATTGATGTTCGCTCAATTCTCAGAG 673
Qy       1378 TTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAG---GACCCG 1434
Db        674 CTTGTGAATGATTACTACAATGGAGGCTTGCCTTCGAATCTGAGTGGTGGGCCTAATCCC 733
Qy       1435 TCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTC 1494
Db        734 AGCCTGGATTATGGACTGAAAGGGGCCGAGATCGCTATGGCTTCTTACACTTCTGAGCTT 793
Qy       1495 GGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCC 1554
Db        794 CTTTACCTGGCAAATCCTGTCACCAGCCATGTACAGAGCGCCGAACAGCATAACCAGGAT 853
Qy       1555 GTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCT 1614
Db        854 GTCAATTCTCTGGGTCTCGTTTCAGCTAGAAAATCTGCCGAGGCCATCGATATTCTGAAG 913
Qy       1615 CTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAG 1674
Db        914 CTGATGCTCTCCACATACCTGACAGCTCTGTGCCAGGCTGTGGATTTAAGGCATCTGGAG 973
Qy       1675 CTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGC 1734
Db        974 GAAAACATGCTGGCCACTGTGAAGCAGATTGTTTCTCAGGTAGCCAAGAAAACCCTGAGC 1033
Qy       1735 ACTGGCCTCGACG 1747
Db       1034 ACAGGGCTCAACG 1046
```

RESULT 13
US-09-252-991A-16475
; Sequence 16475, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16475
; LENGTH: 1548

```
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16475

  Query Match             5.2%;  Score 126.8;  DB 4;  Length 1548;
  Best Local Similarity   47.1%;  Pred. No. 5.1e-16;
  Matches  571;  Conservative    0;  Mismatches  592;  Indels   49;  Gaps    4;

Qy        516 CGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAA 575
Db        351 CGGCGTCGGCGAGGCGCTGGACGACGCCATGGTGCGGCTGGTCATGCTGCTCAAGGTGAA 410

Qy        576 CTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTT 635
Db        411 CAGCCTGGCGCGCGGGTTCTCCGGCATCCGCCGCAAGGTGATCGACGCGCTGATCGCGCT 470

Qy        636 CTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGA 695
Db        471 GATCAACGCCGAGGTCTATCCGCACATCCCGCTGAAAGGCTCGGTGGGTGCGTCCGGCGA 530

Qy        696 CCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCA 755
Db        531 CCTGGCGCCGCTGGCGCACATGTCGCTGGTGCTGATCGGCGAAAGCCGGGCGCGCCATCG 590

Qy        756 CGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGG 815
Db        591 CGGTG------------------AGTGGCTGCCGGCCGCCGAAGCGCTGGCGGTGGCCGG 632

Qy        816 TCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGT 875
Db        633 GCTGGAGCCGCTGACCCTGGCGGCGAAGGAAGGCCTGGCCCTGCTCAATGGCACCCAGGT 692

Qy        876 CTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCA 935
Db        693 GTCCACCGCCTACGCGCTGCGCGGGTTGTTCGAGGCCGAGGACCTGTTCGCCGCCGCCAC 752

Qy        936 GGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTT 995
Db        753 CGTCTGCGGCGGGCTCAGCGTCGAGGCCATGCTCGGTTCGCGGGCGCCGTTCGATGCGCG 812

Qy        996 CATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCAC 1055
Db        813 CATCCATGCCGCGCGCGGCCAGCGTGGGCAGATCGACGTGGCCGCGGCCTATCGCGACCT 872

Qy       1056 GCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGA 1115
Db        873 GCTCACCGCCAGCAGCGAGGTGGCGCGCTCCCATGAGAAGTGCGACAAGGT--------- 923

Qy       1116 CGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCC 1175
Db        924 ---------------CCAGGACCCCTATTCGCTGCGTTGCCAGCCGCAGGTGATGGGCGC 968

Qy       1176 GCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGAC 1235
Db        969 CTGCCTGACCCAGATGCGCCAGGCCGCCGAGGTGCTGGAGATCGAAGCCAACGCGGTGTC 1028

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1029 CGACAACCCGCTGGTATTCGCCGCCGAGGGCGACGTGATCTCCGGCGGCAACTTCCACGC 1088

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1089 CGAACCGGTGGCGATGGCCGCCGACAACCTGGCCCTGGCGTTGGCCGAGATCGGTTCGC- 1147

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1148 -----TGTCGGAACGGCGCATCTCGCTGATGATGGACATGCACATGTCGCAGTTGCCGCC 1202

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1203 GTTCCTGGTGGCCAACGGCGGGGTCAACTCCGGCTTCATGATCGCCCAGGTCACCGCCGC 1262

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1263 CGCGCTGGCCAGCGACAACAAGGCCCTGGCCCATCCGGCCAGCGTCGACAGCCTGCCGAC 1322

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1323 CTCGGCCAACCAGGAAGACCATGTGTCGATGGCGCCGAACGCCGGCAAACGGCTCTGGGC 1382

Qy       1596 GGCCAACGACGTCCTTTCTC-TCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCG 1654
Db       1383 GATGGCCGAGAACGTCCCGCGGGATTCTCGCCGTGGAGTGGCTGGGCGCCTGCCAGGGCC 1442

Qy       1655 TCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTC 1714
Db       1443 TGGACTTCCGCGAGGGCCTGAAGAGTTCGCCGAAGCTGGAGCAGGCTCGCCGCCTGCTGC 1502

Qy       1715 TCCTCCAGCAGC 1726
Db       1503 GCGACAAGGTGC 1514

RESULT 14
US-09-252-991A-16126/c
; Sequence 16126, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 16126
;   LENGTH: 2295
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16126

  Query Match             5.2%;  Score 126.8;  DB 4;  Length 2295;
  Best Local Similarity   47.1%;  Pred. No. 5.4e-16;
  Matches  571;  Conservative    0;  Mismatches  592;  Indels   49;  Gaps    4;

Qy        516 CGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAA 575
Db       1168 CGGCGTCGGCGAGGCGCTGGACGACGCCATGGTGCGGCTGGTCATGCTGCTCAAGGTGAA 1109

Qy        576 CTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTT 635
Db       1108 CAGCCTGGCGCGCGGGTTCTCCGGCATCCGCCGCAAGGTGATCGACGCGCTGATCGCGCT 1049

Qy        636 CTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGA 695
Db       1048 GATCAACGCCGAGGTCTATCCGCACATCCCGCTGAAAGGCTCGGTGGGTGCGTCCGGCGA 989

Qy        696 CCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCA 755
Db        988 CCTGGCGCCGCTGGCGCACATGTCGCTGGTGCTGATCGGCGAAAGCCGGGCGCGCCATCG 929

Qy        756 CGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGG 815
Db        928 CGGTG------------------AGTGGCTGCCGGCCGCCGAAGCGCTGGCGGTGGCCGG 887

Qy        816 TCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGT 875
Db        886 GCTGGAGCCGCTGACCCTGGCGGCGAAGGAAGGCCTGGCCCTGCTCAATGGCACCCAGGT 827

Qy        876 CTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCA 935
Db        826 GTCCACCGCCTACGCGCTGCGCGGGTTGTTCGAGGCCGAGGACCTGTTCGCCGCCGCCAC 767
```

```
Qy        936 GGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTT 995
Db        766 CGTCTGCGGCGGGCTCAGCGTCGAGGCCATGCTCGGTTCGCGGGCGCCGTTCGATGCGCG 707

Qy        996 CATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCAC 1055
Db        706 CATCCATGCCGCGCGCGGCCAGCGTGGGCAGATCGACGTGGCCGCGGCCTATCGCGACCT 647

Qy       1056 GCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGA 1115
Db        646 GCTCACCGCCAGCAGCGAGGTGGCGCGCTCCCATGAGAAGTGCGACAAGGT--------- 596

Qy       1116 CGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCC 1175
Db        595 --------------CCAGGACCCCTATTCGCTGCGTTGCCAGCCGCAGGTGATGGGCGC 551

Qy       1176 GCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGAC 1235
Db        550 CTGCCTGACCCAGATGCGCCAGGCCGCCGAGGTGCTGGAGATCGAAGCCAACGCGGTGTC 491

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db        490 CGACAACCCGCTGGTATTCGCCGCCGAGGGCGACGTGATCTCCGGCGGCAACTTCCACGC 431

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db        430 CGAACCGGTGGCGATGGCCGCCGACAACCTGGCCCTGGCGTTGGCCGAGATCGGTTCGC- 372

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db        371 -----TGTCGGAACGGCGCATCTCGCTGATGATGGACATGCACATGTCGCAGTTGCCGCC 317

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db        316 GTTCCTGGTGGCCAACGGCGGGGTCAACTCCGGCTTCATGATCGCCCAGGTCACCGCCGC 257

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db        256 CGCGCTGGCCAGCGACAACAAGGCCCTGGCCCATCCGGCCAGCGTCGACAGCCTGCCGAC 197

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db        196 CTCGGCCAACCAGGAAGACCATGTGTCGATGGCGCCGAACGCCGGCAAACGGCTCTGGGC 137

Qy       1596 GGCCAACGACGTCCTTTCTC-TCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCG 1654
Db        136 GATGGCCGAGAACGTCCCGCGGGATTCTCGCCGTGGAGTGGCTGGGCGCCTGCCAGGGCC 77

Qy       1655 TCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTC 1714
Db         76 TGGACTTCCGCGAGGGCCTGAAGAGTTCGCCGAAGCTGGAGCAGGCTCGCCGCCTGCTGC 17

Qy       1715 TCCTCCAGCAGC 1726
Db         16 GCGACAAGGTGC 5
```

RESULT 15
US-09-489-039A-6404
; Sequence 6404, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 6404
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6404

```
  Query Match             4.5%;  Score 108.4;  DB 4;  Length 1527;
  Best Local Similarity   48.2%;  Pred. No. 1.9e-12;
  Matches  405;  Conservative    0;  Mismatches  396;  Indels   39;  Gaps    2;

Qy        517 GGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAAC 576
Db        259 GGCGTTGGCGAGCCGCTGGACGATGATCTGGCGCGGCTGATTATGGTGCTGAAGATTAAC 318

Qy        577 TCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTC 636
Db        319 AGCCTGTCCCGCGGTTTCTCCGGGATCCGCCTGAGCGTGATCCAGGCGCTGATTGGCCTG 378

Qy        637 TTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGAC 696
Db        379 GTGAATGCCGGTGTGACGCCGTGGATCCCGGCCAAAGGCTCCGTCGGCGCCTCCGGCGAC 438

Qy        697 CTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCAC 756
Db        439 CTCGCGCCGCTGGCGCATATGTCGCTGACCCTGATCGGAGAAGGTAAAGCGCGGGTACGC 498

Qy        757 GTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGT 816
Db        499 GGCGGCGACTGGC--------------TGCCGGCCACCGAAGCGCTGCGTCAGGTGGGG 543

Qy        817 CTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTC 876
Db        544 CTGGAGCCGATTACCCTGGCGGCGAAAGAGGGGCTGGCGCTACTCAACGGCACCCAGGCC 603

Qy        877 TCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAG 936
Db        604 TCGACCGCTTTTGCCCTGCGCGGTCTGTTCGAAGCCGAGGATCTGTTCGCCTCCGCGGTC 663

Qy        937 GCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTC 996
Db        664 GTCTGCGGGGCGTTAACCACCGAAGCGGCGCTGGGCTCACGTCGTCCGTTTGATGCGCGC 723

Qy        997 ATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACG 1056
Db        724 ATCCATGAAGTGCGCGGCCAGCGCGGGCAGATTGATGCGGCGGCGCTGTATCGCCACCTG 783

Qy       1057 CTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGAC 1116
Db        784 CTGACCGACGATAGCGCGATCTCGCAGTCGCACCATAACTGCAGCAAGGT---------- 833

Qy       1117 GAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCG 1176
Db        834 -------------GCAGGACCCGTACTCCCTGCGCTGCCAGCCGCAGGTGATGGGCGCC 879

Qy       1177 CTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACC 1236
Db        880 TGCCTGACGCAGATCCGCCAGGCGGCTGAGGTGCTGCTGGCTGAAGCCAACGCGGTCTCC 939

Qy       1237 GACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCG 1296
Db        940 GATAACCCGCTGGTCTTTGCCGCCGAAAATGACGTCATCTCCGGCGGTAACTTCCATGCC 999

Qy       1297 TCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTC 1356
Db       1000 GAACCGGTGGCGATGGCCGCGGATAATATCGCGCTGGCGATCGCCGAAATCGGCTCGCTC 1059
```

Search completed: September 12, 2004, 04:26:43
Job time : 190.4 secs

```
OM nucleic - nucleic search, using sw model

Run on:         September 11, 2004, 23:12:54 ; Search time 1570.32 Seconds
                                         (without alignments)
                                         7750.794 Million cell updates/sec

Title:          US-09-939-408A-12
Perfect score:  2419
Sequence:       1 ctcctgcctcactctaaccc..........gagttcaaaaaaaaaaaaaa 2419

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters:      6608766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
                14:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                18:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                19:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2418.6	100.0	2419	9	US-09-939-408A-12	Sequence 12, Appl
2	1826.6	75.5	2741	9	US-09-939-408A-28	Sequence 28, Appl
3	1704	70.4	2163	9	US-09-939-408A-29	Sequence 29, Appl
4	1629	67.3	2475	9	US-09-939-408A-20	Sequence 20, Appl
5	1196.6	49.5	2151	9	US-09-765-873A-9	Sequence 9, Appli
6	1196.6	49.5	2151	15	US-10-138-970A-23	Sequence 23, Appl
7	1196.6	49.5	2151	15	US-10-188-523B-9	Sequence 9, Appli
8	1196.6	49.5	2151	16	US-10-374-366-3	Sequence 3, Appli
9	1196.6	49.5	2439	9	US-09-939-408A-18	Sequence 18, Appl
10	1195	49.4	2151	9	US-09-765-873A-7	Sequence 7, Appli
11	1195	49.4	2151	15	US-10-188-523B-7	Sequence 7, Appli
12	1195	49.4	2151	16	US-10-374-366-1	Sequence 1, Appli
13	1195	49.4	2151	16	US-10-374-366-9	Sequence 9, Appli
14	1195	49.4	2151	16	US-10-374-366-13	Sequence 13, Appl
15	1195	49.4	2151	16	US-10-374-366-23	Sequence 23, Appl
16	1193.4	49.3	2151	16	US-10-374-366-5	Sequence 5, Appli
17	1193.4	49.3	2151	16	US-10-374-366-7	Sequence 7, Appli
18	1193.4	49.3	2151	16	US-10-374-366-15	Sequence 15, Appl
19	1193.4	49.3	2151	16	US-10-374-366-21	Sequence 21, Appl
20	1191.8	49.3	2151	16	US-10-374-366-17	Sequence 17, Appl
21	1190.2	49.2	2151	16	US-10-374-366-11	Sequence 11, Appl
22	1190.2	49.2	2151	16	US-10-374-366-19	Sequence 19, Appl
23	1188.8	49.1	2061	9	US-09-765-873A-31	Sequence 31, Appl
24	1188.8	49.1	2061	15	US-10-188-523B-31	Sequence 31, Appl
25	1136.6	47.0	2787	9	US-09-939-408A-16	Sequence 16, Appl
26	802.4	33.2	2703	16	US-10-439-478-1	Sequence 1, Appli
27	696.8	28.8	2465	17	US-10-439-479-24	Sequence 24, Appl
28	545	22.5	1020	15	US-10-138-970A-3	Sequence 3, Appli
29	424.8	17.6	2554	17	US-10-767-701-15815	Sequence 15815, A
30	423.2	17.5	2539	13	US-10-425-114-26376	Sequence 26376, A
31	408.2	16.9	2325	17	US-10-767-701-15813	Sequence 15813, A
32	402	16.6	2544	17	US-10-437-963-55261	Sequence 55261, A
33	401.6	16.6	2454	13	US-10-425-114-34124	Sequence 34124, A
34	400.2	16.5	2417	13	US-10-425-114-31197	Sequence 31197, A
35	400.2	16.5	2468	13	US-10-425-114-30888	Sequence 30888, A
36	400.2	16.5	2469	13	US-10-425-114-31903	Sequence 31903, A
37	400	16.5	2354	13	US-10-425-114-28070	Sequence 28070, A
38	385.8	15.9	2972	17	US-10-437-963-48311	Sequence 48311, A
39	383	15.8	2774	17	US-10-437-963-38582	Sequence 38582, A
40	382.8	15.8	2329	17	US-10-437-963-41428	Sequence 41428, A
41	382.2	15.8	2079	13	US-10-425-114-1334	Sequence 1334, Ap
42	380.4	15.7	1092	17	US-10-439-479-9	Sequence 9, Appli
43	376.2	15.6	2340	13	US-10-425-114-2857	Sequence 2857, Ap
44	350.8	14.5	2145	17	US-10-437-963-40094	Sequence 40094, A
45	346	14.3	2535	17	US-10-437-963-36111	Sequence 36111, A

ALIGNMENTS

```
RESULT 1
US-09-939-408A-12
; Sequence 12, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION:  Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 2419
;   TYPE: DNA
;   ORGANISM: Rhodotorula graminis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (37)..(2196)
;   NAME/KEY: modified_base
;   LOCATION: (494)
;   OTHER INFORMATION: Other information:  y = t or c
;   NAME/KEY: unsure
;   LOCATION: (493)..(495)
;   OTHER INFORMATION: Other information:  Xaa = Val or Ala
US-09-939-408A-12

  Query Match             100.0%;  Score 2418.6;  DB 9;  Length 2419;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2419;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCCTGCCTCACTCTAACCCGTCACTAGCACTCGCGATGGCCCCTTCCTTGGACTCGCTC 60

Qy         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCACCACGCTCGCCAACGGCTTTACCAACGGCTCGCACGCCGCTCCGACCAAGTCGGCT 120

Qy        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCTCGATGGCCACGCCGCGCACCAG 180

Qy        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTC 240

Qy        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGGGTC 300

Qy        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTCAAG 360

Qy        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGGTTTCGGTGGCTCGGCCGACACG 420

Qy        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTG 480

Qy        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTC 540

Qy        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCC 600

Qy        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATC 660

Qy        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCC 720

Qy        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAG 780

Qy        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCG 840

Qy        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGT 900

Qy        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAG 960

Qy        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCAC 1020

Qy       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCC 1080

Qy       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGC 1140

Qy       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCC 1200

Qy       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAG 1260

Qy       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAG 1320

Qy       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTG 1380

Qy       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTC 1440

Qy       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 AACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCAC 1500

Qy       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAAC 1560

Qy       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTT 1620

Qy       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGAC 1680

Qy       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGC 1740

Qy       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAG 1800

Qy       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 ACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACC 1860

Qy       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCG 1920

Qy       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTC 1980

Qy       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTC 2040

Qy       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 CTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTC 2100

Qy       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGAC 2160

Qy       2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2161 GGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTCGCCTA 2220

Qy        2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2221 GACGCCCGCCTCACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTT 2280

Qy        2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 CCTCCATACACATGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTAT 2340

Qy        2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 CCCGCGTCTCTCGGTCGTCAGTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCG 2400

Qy        2401 AGTTCAAAAAAAAAAAAAA 2419
               |||||||||||||||||||
Db        2401 AGTTCAAAAAAAAAAAAAA 2419


RESULT 2
US-09-939-408A-28
; Sequence 28, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 2741
;   TYPE: DNA
;   ORGANISM: Rhodotorula graminis
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (6)
;   OTHER INFORMATION: Description of modified_base: m = a or c
;   NAME/KEY: exon
;   LOCATION: (2008)..(2586)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (1822)..(1947)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (1587)..(1748)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (1365)..(1529)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (961)..(1295)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (449)..(880)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (1)..(361)
;   OTHER INFORMATION:
;   NAME/KEY: terminator
;   LOCATION: (2587)..(2589)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (1948)..(2007)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (1749)..(1821)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (1530)..(1586)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (1296)..(1364)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (881)..(960)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (362)..(448)
;   OTHER INFORMATION:
US-09-939-408A-28

  Query Match             75.5%;  Score 1826.6;  DB 9;  Length 2741;
  Best Local Similarity   84.3%;  Pred. No. 0;
  Matches 2312;  Conservative    2;  Mismatches    1;  Indels  426;  Gaps    6;

Qy          37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
               ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGGCMCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 60

Qy          97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 156
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 120

Qy         157 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 216
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 180

Qy         217 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 276
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 240

Qy         277 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 336
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 300

Qy         337 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCAC- 395
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db         301 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACG 360

Qy         396 ------------------------------------------------------------ 395
                                                                           
Db         361 GTGCGTTCCGAGACGAGAGGCGGAAATCTCGGGATGCCGCAGCGCTGAACGCTGACACTC 420

Qy         396 --------------------------GGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 429
                                         ||||||||||||||||||||||||||||||||||
Db         421 GCTTGGACGGCTGCCGCGGTCTTGCAGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAG 480

Qy         430 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 489
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACG 540

Qy         490 TCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 549
               ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 TCCGTCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTC 600

Qy         550 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 609
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 660

Qy         610 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 669
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 720

Qy         670 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 729
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 780
```

```
Qy    730 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 840

Qy    790 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTC--------------------- 828
          |||||||||||||||||||||||||||||||||||||||
Db    841 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGGTACGTCGCGAGTCCTGACT 900

Qy    829 -----------------------------------------------------------G 829
                                                                     |
Db    901 GCAGTGAGCTGTTCGAGAGTCTCCCAGTTTGCTGACTGCCCTTTGTTCATGCGATTGCAG 960

Qy    830 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 889
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGG 1020

Qy    890 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTC 1080

Qy    950 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCT 1140

Qy   1010 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCT 1069
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCT 1200

Qy   1070 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTC 1260

Qy   1130 GCCAGGACCGCTACCCGCTCCGCACGTCGCCT---------------------------- 1161
          ||||||||||||||||||||||||||||||||
Db   1261 GCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGGTTCGTCCCCTCTCTCCCCTTCC 1320

Qy   1162 -----------------------------------------CAGTTCCTCGGCCCGCTCG 1180
                                                   |||||||||||||||||||
Db   1321 CTCCGTCCGACCGGCGCGTCGAGACTTACGTTTTGCGTATCCAGTTCCTCGGCCCGCTCG 1380

Qy   1181 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACA 1440

Qy   1241 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGG 1500

Qy   1301 CTGTCTCGATTTCGATGGAGAAGAC----------------------------------- 1325
          |||||||||||||||||||||||||
Db   1501 CTGTCTCGATTTCGATGGAGAAGACCAGGTGCGTCTCTCGCTGCCTTCGTACTCCGATCT 1560

Qy   1326 ----------------------CAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1363
                                ||||||||||||||||||||||||||||||||||||||
Db   1561 TGTGCTGAATGTTCTTCTCCTGCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCA 1620

Qy   1364 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG 1680

Qy   1424 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1740

Qy   1484 CTTCGG------------------------------------------------------ 1489
          ||||||
Db   1741 CTTCGGAGGTGAGCCGTCGACGTTCTCCGCCGTCGCTCGTCCCCTTCAGCGCACCCAGGC 1800

Qy   1490 -------------------AGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1530
                             |||||||||||||||||||||||||||||||||||||||||
Db   1801 TGACTTCCTTTCCCTCTGTAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAG 1860

Qy   1531 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACT 1920

Qy   1591 GCCGAGGCCAACGACGTCCTTTCTCTC--------------------------------- 1617
          |||||||||||||||||||||||||||
Db   1921 GCCGAGGCCAACGACGTCCTTTCTCTCGTGCGTTCGTGTCGCAATGAGTCCCGACGCAAT 1980

Qy   1618 ---------------------------CTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 1650
                                     |||||||||||||||||||||||||||||||||
Db   1981 AGCGACTGACTGCGCGATCCTGAGCAGCTTCTCGCCTCGCACCTGTACTGCACGCTCCAG 2040

Qy   1651 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 1710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 GCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCG 2100

Qy   1711 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 1770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTC 2160

Qy   1771 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 1830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 AAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGG 2220

Qy   1831 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 1890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCT 2280

Qy   1891 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATC 2340

Qy   1951 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 TCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCG 2400

Qy   2011 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGC 2460

Qy   2071 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAAC 2520

Qy   2131 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2190
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATG 2580

Qy   2191 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CTCGCGTAAGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTCACCCCAAGACCAGCTTTT 2640

Qy   2251 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGCCTTACTCTCTCG 2310
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   2641 CGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACATGTCGTCTTACTCTCTCG 2700

Qy   2311 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2351
          |||||||||||||||||||||||||||||||||||||||||
Db   2701 CCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCT 2741
```

RESULT 3
US-09-939-408A-29
; Sequence 29, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A

```
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 2163
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   Preferred
;   OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
;   NAME/KEY: CDS
;   LOCATION: (1)..(2163)
;   OTHER INFORMATION:
US-09-939-408A-29

  Query Match             70.4%;  Score 1704;  DB 9;  Length 2163;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches 1817;  Conservative  109;  Mismatches  236;  Indels    0;  Gaps    0;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
Db          1 ATGGCCCCCTCCBTCGACTCGATCGCGACCTCGBTCGCCAACGGCBTCDCVAACGGVTHG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGC 156
Db         61 CACGCCGCTCCGHCCAAGYCGSCWACGGGCGCCACGTCCACHCTCMGMCGGCCGBCCDGG 120

Qy        157 CTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGAC 216
Db        121 CTCGCTCCTCCCGCCACCCAGVHGACGCAGCTCGACATCGTBGAGVAGATCCTCGCCGAC 180

Qy        217 CCCACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGC 276
Db        181 CCCACCGACGACGWSVTCGAACTCGACGGGTACACCCTCACCCTCGGHGACGTCGTCGGC 240

Qy        277 GCCGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTC 336
Db        241 GCCGCBCGCAAGGGCCGCHCBGTCCGCGTCCAGACAGMCGACGAGATCCGCGCAAAGATC 300

Qy        337 GACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACG 396
Db        301 GACAAVAGBGTCGAGTTCCTCCGBDCBCAGCTCBACAACAGBGTCTACGGHGTCACGACT 360

Qy        397 GGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTC 456
Db        361 GGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCVATCTCGCTCCAGAAGGCBCTC 420

Qy        457 ATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGC 516
Db        421 CTCGAGCACCAGCTCTGCGGTGTBCTCCCBACGTCGDTCGABTCCTTCVGCCTCGGHCGC 480

Qy        517 GGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAAC 576
Db        481 GGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGCGTCAAC 540

Qy        577 TCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTC 636
Db        541 TCGCTCACBCGCGGCCACTCGGCBGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTC 600

Qy        637 TTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGAC 696
Db        601 CTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCGGGCGAC 660

Qy        697 CTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCAC 756
Db        661 CTCTCCCCBCTCTCBTACATCGCCGCCGCCATCACCGGTCACCCGGACDBCAAGGTBCAC 720

Qy        757 GTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGT 816
Db        721 GTYKTSCACGAGGGCAMSGAGAAGATCATGTHCGCCCGCGAGGCGATCGCGCTCTTBGGT 780

Qy        817 CTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTC 876
Db        781 CTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACGGCCGTC 840

Qy        877 TCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAG 936
Db        841 TCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAG 900

Qy        937 GCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTC 996
Db        901 GCGCTCACGGCTCTBACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCACCCVTTC 960

Qy        997 ATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACG 1056
Db        961 CTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATCCGCACG 1020

Qy       1057 CTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGAC 1116
Db       1021 CTCCTCGAGGGCAGCHVGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAGGACGAC 1080

Qy       1117 GAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCG 1176
Db       1081 GAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTCGGCCCG 1140

Qy       1177 CTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACC 1236
Db       1141 CTCGTCAGCGACATGATTCACGCCCACGCBGTCCTCTCGCTCGAGGCCGAGTCGACGACC 1200

Qy       1237 GACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCG 1296
Db       1201 GACAACCCGCTCATCGACGTCGAGAACAAGAHGACCCACCACGGCGGCAACTTCCAGGCG 1260

Qy       1297 TCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTC 1356
Db       1261 DCCGCTGTCGCVAACACGATGGAGAAGACTCGCCTCGCVCTCGCCCTGATCGGCAAGCTC 1320

Qy       1357 AACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGC 1416
Db       1321 AACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTBCCBTCCTGC 1380

Qy       1417 CTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCT 1476
Db       1381 CTCGCTGCCGAGGACCCBTCGCTCTCCTATCACTGCAAGGGCCTCGACATTGCCGCBGCB 1440

Qy       1477 GCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCA 1536
Db       1441 GCBTACACTTCGGAGCTCGGHCACCTTGCCAACCCGGTBACGACCCACGTCCAGCCGGCH 1500

Qy       1537 GAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCCGCGCGCCGCACTGCCGAG 1596
Db       1501 GAGATGGGCAACCAGGCCGTCAACTCGCTCGCBCTCATCTCGGCBCGCCGCACBGCCGAG 1560

Qy       1597 GCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTC 1656
Db       1561 GCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTBCTCCAGGCCGTC 1620

Qy       1657 GACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTC 1716
Db       1621 GACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGVTBVTCVCBDCGCTC 1680

Qy       1717 CTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCACTTGCGCTCGAGGTCAAGAAG 1776
Db       1681 HTCVAGCAGCACTTTGGCDCYGCCCTCGACGGCWACGAACTCGHGGACAAGGTCAACAAG 1740

Qy       1777 GCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGGCACGAC 1836
Db       1741 DCGCTCDACAAGCGBCTCGAGCAGACCAACTCGTACGACCTCGAGCCGCGCTGGCACGAC 1800

Qy       1837 GCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAAC 1896
Db       1801 GCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCBYCTGCCAAG 1860
```

```
Qy       1897 GTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTC 1956
Db       1861 GTCTCGCTCGCBGCCGTCAACGCCTGGAAGGTCGCCTCCGCCGAGAAGGCCATCTCGCTC 1920

Qy       1957 ACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCA 2016
Db       1921 ACGCGCBAVGTCCGCGACHCCTTCTGGBCGGCBCCGTCGTCGTCGTCGCCCGCGCTCDCG 1980

Qy       2017 TACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAG 2076
Db       1981 TACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTCCGCGAGGAGCTCGGCGTCAAG 2040

Qy       2077 GCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCG 2136
Db       2041 GCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGGTGACGATCGGCACCAACGTCTCC 2100

Qy       2137 CGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCG 2196
Db       2101 CGCATCTACGAGGCCATCAAGDVCGGCHGCATCAACCACGTCCTCGTCAAGATGCTCGCD 2160

Qy       2197 TA 2198
Db       2161 TA 2162
```

RESULT 4
US-09-939-408A-20
; Sequence 20, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2475
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_difference
; LOCATION: (13)..(2475)
; OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
; OTHER INFORMATION: Sequence of SEQ ID NOs: 12, 16, and 18
US-09-939-408A-20

```
  Query Match             67.3%;  Score 1629;  DB 9;  Length 2475;
  Best Local Similarity   78.0%;  Pred. No. 0;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347
```

```
Db    1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy    1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407
Db    1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy    1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467
Db    1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy    1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527
Db    1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy    1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587
Db    1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy    1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647
Db    1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy    1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707
Db    1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy    1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC-----------AACGCA 1755
Db    1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy    1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815
Db    1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy    1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875
Db    1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy    1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933
Db    1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy    1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993
Db    1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy    1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053
Db    1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy    2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113
Db    2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy    2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173
Db    2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy    2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232
Db    2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy    2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292
Db    2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy    2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352
Db    2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy    2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411
Db    2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy    2412 AAAAAAAA 2419
Db    2401 NNNANAAA 2408
```

RESULT 5
US-09-765-873A-9
; Sequence 9, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-9

```
  Query Match             49.5%;  Score 1196.6;  DB 9;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 2.4e-282;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy      99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db      45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy     159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db     105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy     219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db     165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy     279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db     225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy     339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db     285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy     399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db     345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy     459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db     405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy     519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db     465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy     579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db     525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy     639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
```

```
Db      585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy      699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db      645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy      759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db      705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy      819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db      765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy      879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db      825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy      939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db      885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy      999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db      945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy     1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db     1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy     1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db     1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy     1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db     1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy     1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db     1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy     1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db     1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy     1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db     1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy     1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db     1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy     1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db     1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy     1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db     1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy     1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db     1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy     1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db     1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy     1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db     1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy     1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db     1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198
Db     2136 CAAGATGCTCGCTTA 2150
```

RESULT 6
US-10-138-970A-23
; Sequence 23, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: artificial sequence
;   FEATURE:
;   OTHER INFORMATION: mutant from Rhodotorula glutinis
US-10-138-970A-23

```
  Query Match            49.5%;  Score 1196.6;  DB 15;  Length 2151;
  Best Local Similarity  74.0%;  Pred. No. 2.4e-282;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy       99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db       45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy      159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db      105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy      219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db      165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224
```

```
Qy    279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db    225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy    339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db    285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy    399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db    345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy    459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db    405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy    519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db    465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy    579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db    525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy    639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db    585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy    699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db    645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy    759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db    705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy    819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db    765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy    879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db    825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy    939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db    885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy    999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db    945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy   1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db   1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy   1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db   1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy   1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db   1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy   1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db   1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy   1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db   1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy   1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db   1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy   1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db   1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy   1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db   1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy   1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db   1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy   1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db   1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy   1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db   1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy   1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db   1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy   1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db   1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy   1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db   1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy   1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db   1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy   1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db   1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy   2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db   1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy   2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db   2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy   2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db   2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy   2184 CAAGATGCTCGCGTA 2198
Db   2136 CAAGATGCTCGCTTA 2150
```

RESULT 7
US-10-188-523B-9
; Sequence 9, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27

```
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis mutant
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2151)
;   OTHER INFORMATION:
US-10-188-523B-9

  Query Match             49.5%;  Score 1196.6;  DB 15;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 2.4e-282;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955
```

```
Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 8
US-10-374-366-3
; Sequence 3, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-3

```
  Query Match             49.5%;  Score 1196.6;  DB 16;  Length 2151;
  Best Local Similarity   74.0%;  Pred. No. 2.4e-282;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCCGCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604
```

```
Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 9
US-09-939-408A-18
; Sequence 18, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 2439
;   TYPE: DNA
;   ORGANISM: Rhodotorula toruloides
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2148)
US-09-939-408A-18

```
  Query Match             49.5%;  Score 1196.6;  DB 9;  Length 2439;
  Best Local Similarity   74.0%;  Pred. No. 2.4e-282;
  Matches 1565;  Conservative    1;  Mismatches  525;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
```

```
Db        1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy        1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db        1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy        1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db        1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy        1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db        1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy        1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db        1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy        1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db        1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy        1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db        1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy        1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db        1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy        1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db        1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy        1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db        1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy        1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db        1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy        1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db        1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy        1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db        1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy        1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db        1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy        2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db        1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy        2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db        2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy        2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db        2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy        2184 CAAGATGCTCGCGTA 2198
Db        2136 CAAGATGCTCGCTTA 2150
```

RESULT 10

```
US-09-765-873A-7
; Sequence 7, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2151)
US-09-765-873A-7

  Query Match             49.4%;  Score 1195;  DB 9;  Length 2151;
  Best Local Similarity   73.9%;  Pred. No. 5.9e-282;
  Matches 1564;  Conservative    1;  Mismatches  526;  Indels   24;  Gaps    3;

Qy          99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db          45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy         159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db         105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy         219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db         165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy         279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db         225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy         339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db         285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy         399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db         345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy         459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db         405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy         519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db         465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy         579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db         525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy         639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db         585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy         699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db         645 GTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy         759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
```

```
Db    705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy    819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db    765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy    879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db    825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTCTCGCAGTC 884

Qy    939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db    885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy    999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db    945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy   1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db   1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy   1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db   1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy   1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db   1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy   1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db   1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy   1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db   1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy   1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db   1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy   1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db   1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy   1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db   1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy   1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db   1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy   1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db   1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy   1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db   1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy   1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db   1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy   1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db   1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy   1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db   1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy   1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db   1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy   1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db   1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy   2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db   1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy   2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db   2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy   2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db   2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy   2184 CAAGATGCTCGCGTA 2198
Db   2136 CAAGATGCTCGCTTA 2150
```

RESULT 11
US-10-188-523B-7
; Sequence 7, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2151)
;   OTHER INFORMATION:
US-10-188-523B-7

Query Match 49.4%; Score 1195; DB 15; Length 2151;
Best Local Similarity 73.9%; Pred. No. 5.9e-282;
Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

```
Qy     99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db     45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy    159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db    105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy    219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db    165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy    279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db    225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284
```

```
Qy    339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db    285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy    399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db    345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy    459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db    405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy    519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db    465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy    579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db    525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy    639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db    585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy    699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db    645 GTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy    759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db    705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy    819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db    765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy    879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db    825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTCTCGCAGTC 884

Qy    939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db    885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy    999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db    945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy   1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db   1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy   1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db   1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy   1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db   1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy   1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db   1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy   1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db   1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy   1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db   1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy   1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db   1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy   1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db   1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy   1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db   1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy   1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db   1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy   1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db   1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy   1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db   1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy   1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db   1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy   1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db   1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy   1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db   1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy   1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db   1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy   2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db   1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy   2064 GCTCGGCGTGCAGGCGCGCCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db   2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy   2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db   2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy   2184 CAAGATGCTCGCGTA 2198
Db   2136 CAAGATGCTCGCTTA 2150
```

```
RESULT 12
US-10-374-366-1
; Sequence 1, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 2151
```

```
;  TYPE: DNA
;  ORGANISM: Rhodosporidium glutinis
US-10-374-366-1

  Query Match             49.4%;  Score 1195;  DB 16;  Length 2151;
  Best Local Similarity   73.9%;  Pred. No. 5.9e-282;
  Matches 1564;  Conservative    1;  Mismatches  526;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 GTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db       1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy       1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db       1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy       1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db       1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy       1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db       1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy       1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db       1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy       1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db       1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy       1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA------------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
```

```
)b        2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

)y        2184 CAAGATGCTCGCGTA 2198

)b        2136 CAAGATGCTCGCTTA 2150


RESULT 13
JS-10-374-366-9
; Sequence 9, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodosporidium glutinis
JS-10-374-366-9

  Query Match           49.4%;  Score 1195;  DB 16;  Length 2151;
  Best Local Similarity  73.9%;  Pred. No. 5.9e-282;
  Matches 1564;  Conservative    1;  Mismatches  526;  Indels   24;  Gaps    3;

)y          99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158

)b          45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

)y         159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218

)b         105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

)y         219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278

)b         165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

)y         279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338

)b         225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

)y         339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398

)b         285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

)y         399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458

)b         345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

)y         459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518

)b         405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

)y         519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578

Db         465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy         579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638

Db         525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy         639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698

Db         585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy         699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758

Db         645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy         759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818

Db         705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy         819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878

Db         765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy         879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938

Db         825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy         939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998

Db         885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy         999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058

Db         945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy        1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118

Db        1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy        1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178

Db        1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy        1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235

Db        1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy        1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295

Db        1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy        1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355

Db        1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy        1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415

Db        1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy        1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475

Db        1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy        1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535

Db        1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy        1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595

Db        1485 TGAGATGGCGAACCAGGCGGGCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy        1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655

Db        1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy        1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715

Db        1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy        1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763

Db        1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy        1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
```

```
Db     1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy     1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db     1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy     1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db     1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy     1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db     1896 GGCCATCTCGCTCACCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy     2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db     1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy     2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db     2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy     2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db     2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy     2184 CAAGATGCTCGCGTA 2198
Db     2136 CAAGATGCTCGCTTA 2150
```

RESULT 14
US-10-374-366-13
; Sequence 13, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodosporidium glutinis
US-10-374-366-13

```
  Query Match             49.4%;  Score 1195;  DB 16;  Length 2151;
  Best Local Similarity   73.9%;  Pred. No. 5.9e-282;
  Matches 1564;  Conservative    1;  Mismatches  526;  Indels   24;  Gaps    3;

Qy       99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db       45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy      159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db      105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy      219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db      165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy      279 CGCCCGCAAGGGGCGCAGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db      225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy      339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db      285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy      399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db      345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy      459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db      405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy      519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db      465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy      579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db      525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy      639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db      585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy      699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db      645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy      759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db      705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy      819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db      765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy      879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db      825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy      939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db      885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy      999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db      945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy     1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
Db     1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAGGACGACGA 1064

Qy     1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db     1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy     1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db     1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy     1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db     1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCGCGGCGGCAATTTCCAGGC 1244

Qy     1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db     1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy     1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db     1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy     1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
```

```
Db       1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy       1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db       1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy       1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db       1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy       1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db       1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy       1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db       1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy       1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db       1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy       1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db       1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy       1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db       1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy       1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db       1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy       1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db       1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy       2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db       1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy       2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db       2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy       2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db       2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy       2184 CAAGATGCTCGCGTA 2198
Db       2136 CAAGATGCTCGCTTA 2150
```

RESULT 15
US-10-374-366-23
; Sequence 23, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-23

```
  Query Match             49.4%;  Score 1195;  DB 16;  Length 2151;
  Best Local Similarity   73.9%;  Pred. No. 5.9e-282;
  Matches 1564;  Conservative    1;  Mismatches  526;  Indels   24;  Gaps    3;

Qy         99 CGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGCCCGGCCT 158
Db         45 CGTCGCATCCGCAAAGCAGGCTGTCAATGGCGCCTCGACCAACCTCGCAGTCGCAGGCTC 104

Qy        159 CGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCC 218
Db        105 GCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCGCGCC 164

Qy        219 CACCGACGACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTGTCGGCGC 278
Db        165 GACCGACTCGACGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCGTCTCGGC 224

Qy        279 CGCCCGCAAGGGGCGCAGGGTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGA 338
Db        225 CGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAGCGACGAGATCCGCTCAAAGATTGA 284

Qy        339 CAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTCACCACGGG 398
Db        285 CAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTCACGACTGG 344

Qy        399 TTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCAT 458
Db        345 ATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAGGCTCTCCT 404

Qy        459 CGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGG 518
Db        405 CGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTCGGCCGCGG 464

Qy        519 CCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTC 578
Db        465 TCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGCGTCAACAG 524

Qy        579 GCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTT 638
Db        525 CTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACCAACTTCCT 584

Qy        639 GAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCT 698
Db        585 CAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCGGGCGACCT 644

Qy        699 CAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGT 758
Db        645 CTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAGGTGCACGT 704

Qy        759 TTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCT 818
Db        705 CGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCT 764

Qy        819 CGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTC 878
Db        765 CGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACCGCCGTCTC 824

Qy        879 CGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGC 938
Db        825 AGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGTC 884

Qy        939 CTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCAT 998
Db        885 GCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCACCCCTTCCT 944

Qy        999 CCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCT 1058
Db        945 TCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATCCGCAAGCT 1004

Qy       1059 CCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGA 1118
```

```
Db    1005 CCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAGGGACGACGA 1064

Qy    1119 GGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCT 1178
Db    1065 GGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTCGGCCCGCT 1124

Qy    1179 CGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACA---CGACGAC 1235
Db    1125 CGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAGTCGACGAC 1184

Qy    1236 CGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGC 1295
Db    1185 CGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAATTTCCAGGC 1244

Qy    1296 GTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCT 1355
Db    1245 TGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATCGGCAAGCT 1304

Qy    1356 CAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTG 1415
Db    1305 CAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTCCCCTCCTG 1364

Qy    1416 CCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGC 1475
Db    1365 CCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATCGCCGCTGC 1424

Qy    1476 TGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGC 1535
Db    1425 GGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTCCAGCCGGC 1484

Qy    1536 AGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGA 1595
Db    1485 TGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGCACGACCGA 1544

Qy    1596 GGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGT 1655
Db    1545 GTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTCCAAGCCAT 1604

Qy    1656 CGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCT 1715
Db    1605 CGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATCGTCTCGCT 1664

Qy    1716 CCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCA-----------ACGCACTTGCGCT 1763
Db    1665 CATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAGCTCGTCGA 1724

Qy    1764 CGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCC 1823
Db    1725 GAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGACCTCGTCCC 1784

Qy    1824 GCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTCTCGTCCTC 1883
Db    1785 GCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTCTCGTCGAC 1844

Qy    1884 GCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAA 1943
Db    1845 GTC---------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCGCCGCCGAGTC 1895

Qy    1944 GGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGC 2003
Db    1896 GGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGTCGACCTCGTC 1955

Qy    2004 GCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGA 2063
Db    1956 GCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCGTCCGCGAGGA 2015

Qy    2064 GCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGG 2123
Db    2016 GCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGGTGACGATCGG 2075

Qy    2124 GAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGT 2183
Db    2076 CTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACAACGTCCTCCT 2135

Qy    2184 CAAGATGCTCGCGTA 2198
Db    2136 CAAGATGCTCGCTTA 2150
```

Search completed: September 12, 2004, 12:22:08
Job time : 1583.32 secs

```
OM nucleic - nucleic search, using sw model

Run on:          September 11, 2004, 13:17:03 ; Search time 9123.4 Seconds
                                            (without alignments)
                                            7917.733 Million cell updates/sec

Title:           US-09-939-408A-12
Perfect score:   2419
Sequence:        1 ctcctgcctcactctaaccc..........gagttcaaaaaaaaaaaaa 2419

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:      55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        EST:*
                 1:  em_estba:*
                 2:  em_esthum:*
                 3:  em_estin:*
                 4:  em_estmu:*
                 5:  em_estov:*
                 6:  em_estpl:*
                 7:  em_estro:*
                 8:  em_htc:*
                 9:  gb_est1:*
                 10:  gb_est2:*
                 11:  gb_htc:*
                 12:  gb_est3:*
                 13:  gb_est4:*
                 14:  gb_est5:*
                 15:  em_estfun:*
                 16:  em_estom:*
                 17:  em_gss_hum:*
                 18:  em_gss_inv:*
                 19:  em_gss_pln:*
                 20:  em_gss_vrt:*
                 21:  em_gss_fun:*
                 22:  em_gss_mam:*
                 23:  em_gss_mus:*
                 24:  em_gss_pro:*
                 25:  em_gss_rod:*
                 26:  em_gss_phg:*
                 27:  em_gss_vrl:*
                 28:  gb_gss1:*
                 29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	400.2	16.5	2598	11	AY103647	AY103647 Zea mays
	2	386.2	16.0	2299	11	AY106831	AY106831 Zea mays
	3	289.2	12.0	1864	11	AY104679	AY104679 Zea mays
	4	268	11.1	859	14	CB632830	CB632830 OSIIEb11F
c	5	266.8	11.0	866	14	CB645202	CB645202 OSJNEb06M
c	6	266.6	11.0	860	14	CB676595	CB676595 OSJNEe12O
c	7	265.6	11.0	889	14	CB648876	CB648876 OSJNEb12H
c	8	265.4	11.0	864	14	CB649712	CB649712 OSJNEb13M
c	9	265.2	11.0	867	14	CB643557	CB643557 OSJNEb04F
c	10	265.2	11.0	867	14	CB651503	CB651503 OSJNEb16H
c	11	265	11.0	860	14	CB658123	CB658123 OSJNEc13O
c	12	263.8	10.9	891	14	CB651685	CB651685 OSJNEb16L
c	13	263.8	10.9	893	14	CB648470	CB648470 OSJNEb11N
c	14	263.6	10.9	868	14	CB673619	CB673619 OSJNEe08D
c	15	263	10.9	858	14	CB647704	CB647704 OSJNEb10L
c	16	262.6	10.9	857	14	CB677847	CB677847 OSJNEe15D
c	17	262.6	10.9	868	14	CB673025	CB673025 OSJNEe07F
c	18	262.2	10.8	855	14	CB649386	CB649386 OSJNEb13E
c	19	262.2	10.8	907	14	CB643494	CB643494 OSJNEb04D
c	20	262	10.8	862	14	CB649699	CB649699 OSJNEb13M
	21	262	10.8	909	14	CB629973	CB629973 OSIIEb06K
c	22	261.8	10.8	864	14	CB650939	CB650939 OSJNEb15K
c	23	261.6	10.8	855	14	CB675774	CB675774 OSJNEe11L
c	24	261.4	10.8	851	14	CB666100	CB666100 OSJNEd12N
	25	260.4	10.8	856	14	CB654478	CB654478 OSJNEc07A
c	26	258.2	10.7	853	14	CB644159	CB644159 OSJNEb05D
	27	257.4	10.6	837	29	CG209714	CG209714 OGWAE58TH
c	28	255.8	10.6	846	14	CB644856	CB644856 OSJNEb06E
c	29	255.4	10.6	861	14	CB650414	CB650414 OSJNEb14N
c	30	255.4	10.6	864	14	CB669223	CB669223 OSJNEe01I
c	31	255	10.5	846	14	CB641706	CB641706 OSJNEb01H
c	32	254.8	10.5	840	14	CB651026	CB651026 OSJNEb15N
c	33	254.8	10.5	845	14	CB650590	CB650590 OSJNEb15B
c	34	254.8	10.5	845	14	CB651106	CB651106 OSJNEb15P
c	35	254.8	10.5	845	14	CB651730	CB651730 OSJNEb16N
c	36	254.8	10.5	845	14	CB665372	CB665372 OSJNEd11L
c	37	254.8	10.5	847	14	CB644050	CB644050 OSJNEb05B
	38	254.8	10.5	928	14	CB672850	CB672850 OSJNEe07B
c	39	254.6	10.5	841	14	CB676570	CB676570 OSJNEe12O
c	40	254.6	10.5	862	14	CB646327	CB646327 OSJNEb08L
c	41	253.8	10.5	845	14	CB644800	CB644800 OSJNEb06C
c	42	253.8	10.5	852	14	CB666405	CB666405 OSJNEd13E
c	43	253.8	10.5	867	14	CB651444	CB651444 OSJNEb16G
c	44	253.2	10.5	845	14	CB675728	CB675728 OSJNEe11K
c	45	253.2	10.5	865	14	CB642929	CB642929 OSJNEb03F

ALIGNMENTS

```
RESULT 1
AY103647
LOCUS       AY103647                2598 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO142084 mRNA sequence.
ACCESSION   AY103647
VERSION     AY103647.1  GI:21206725
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2598)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2598)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
```

```
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .2598
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638378"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match             16.5%;  Score 400.2;  DB 11;  Length 2598;
  Best Local Similarity   53.4%;  Pred. No. 6.3e-45;
  Matches 1095;  Conservative    0;  Mismatches  903;  Indels   53;  Gaps   10;

Qy        184 CAGCTCGAGATCGTGCAGGAGCTCCTCAGCGACCCCACCGACGACGTCGTCGAGCTCAGC 243
Db        185 CACCTGGACGAGGTGAAGCGCATGGTGGCGCAGGCCCGGCAGCCCGTGGTCAAGATCGAG 244

Qy        244 GGGTACAGCCTCACCGTCCG------TGACGTTGTCGGCGCCGCCCGCAAGGGGCGCAGG 297
Db        245 GGTTCCACCCTCCGCGTCGGCCAGGTGGCCGCCGTCGCCTCCGCCAAGGACGCGTCCGGC 304

Qy        298 GTCCGCGTCCAGAACGACGACGAGATCCGCGCACGCGTCGACAAGAGCGTCGACTTCCTC 357
Db        305 GTCGCCGTCGAGCTCGACGAGGAGGCCCGCCCCCGCGTCAAGGCCAGCAGCGAGTGGATC 364

Qy        358 AAGGCCCAGCTTCAGAACTCGG------TCTACGGAGTCACCACGGGTTTCGGTGGCTCG 411
Db        365 CTCGACTGCATCGCCCACGGCGGCGACATCTACGGCGTCACCACCGGCTTCGGCGGCACC 424

Qy        412 GCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAGGCGCTCATCGAGCACCAGCTC 471
Db        425 TCCCACCGCCGCACCAAGGACGGGCCCGCGCTCCAGGTCGAGCTGCTCAGGCATCTCAAC 484

Qy        472 TGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTCGGACGCGGCCTCGAGAACACG 531
Db        485 GCCGGA-----------------------ATCTTCGGCACCGGCAGCGACGGGCACACG 520

Qy        532 CTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGC 591
Db        521 CTGCCGTCGGAGGTCACCCGCGCGGCGATGCTGGTGCGCATCAACACCCTCCTCCAGGGC 580

Qy        592 CACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATC 651
Db        581 TACTCCGGCATCCGCTTCGAGATCCTCGAGGCCATCACGAAGCTGCTCAACACCGGTGTC 640

Qy        652 ACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCG 711
Db        641 AGCCCCTGCCTGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTGGTCCCGCTCTCC 700

Qy        712 TACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGA 771
Db        701 TACATCGCCGGCCTCATCACGGGCCGCCCCAACGCGCAGGCCGTCACCGTCGACGGGAGG 760

Qy        772 ACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTC 831
Db        761 AAGGTGGACGCCGCCGAGGCGTTCAAG---ATCGCCGGCATCGAGGGCGGCTTCTTCAAG 817

Qy        832 CTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCG 891
Db        818 CTCAACCCCAAGGAGGGCCTCGCCATCGTCAACGGCACGTCCGTGGGCTCCGCGCTCGCG 877

Qy        892 ACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTC 951
Db        878 GCCACCGTGATGTACGACGCCAACGTCCTGGCCGTCCTGTCGGAGGTCCTGTCCGCCGTC 937

Qy        952 ACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGC 1011
Db        938 TTCTGCGAGGTCATGAACGGCAAGCCCGAGTACACGGACCACCTGACCCACAAGCTGAAG 997

Qy       1012 CGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCG 1071
Db        998 CAC---CACCCGGGGTCCATCGAGGCCGCGGCCATCATGGAGCACATCCTGGATGGCAGC 1054

Qy       1072 TCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGC 1131
Db       1055 TCCTT---CATGAAGCAGGCCAAGAAGTTGAACGAGCTGGACCCGCTGCTGAAGCCCAAG 1111

Qy       1132 CAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATG 1191
Db       1112 CAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTGGGCCCCCAGATCGAGGTCATC 1171

Qy       1192 ATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTC 1251
Db       1172 CGCGCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTCATC 1231

Qy       1252 GACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATT 1311
Db       1232 GACGTCCACCGCGGCAAGGCGCTGCACGGCGGCAACTTCCAGGGCACCCCCATCGGCGTG 1291

Qy       1312 TCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGC 1371
Db       1292 TCCATGGACAACGCCCGCCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTC 1351

Qy       1372 ACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAG 1428
Db       1352 TCCGAGCTCGTCAACGAGTTCTACAACAACGGGCTCACCTCCAACCTGGCCGGCAGCCGC 1411

Qy       1429 GACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCG 1488
Db       1412 AACCCCAGCCTGGACTACGGCTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCC 1471

Qy       1489 GAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAAC 1548
Db       1472 GAGCTCCAGTACCTGGGCAACCCCATCACCAACCACGTGCAGAGCGCGGAGCAGCACAAC 1531

Qy       1549 CAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTC 1608
Db       1532 CAGGACGTGAACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCCGAGGCGATCGACATC 1591

Qy       1609 CTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCG 1668
Db       1592 CTGAAGCTCATGTCGTCCACCTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCAC 1651

Qy       1669 ATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTCCCGACTCTCCTCCAGCAGCAC 1728
Db       1652 CTCGAGGAGAACATCAAGGCGTCGGTGAAGAACACCGTGACCCAGGTGGCCAAGAAGGTG 1711

Qy       1729 CTCGGCACTGGCCTCGACGTCAA---CGCACTTGCGCTCGAGGTCAAGAAGGCGCTCAAC 1785
Db       1712 CTGACCATGAACCCCTCGGGCGAGCTCTCCAGCGCCCGCTTCAGCGAGAAGGAGCTGATC 1771

Qy       1786 AAGCGTCTCGAGCAGACGACGACGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCG 1845
Db       1772 AGCGCCATCGACCGCGAGGCCGTGTTCACGTACGCGGAGGACGCGGCCAGCGGCAGCCTG 1831

Qy       1846 TACGCGA-CCGGCACCGTCGTCGAGCTCCTCTCGTCCTCGCCCTCTGCCAACGTCACCCT 1904
Db       1832 CCGCTGATGCAGAAGCTGCGCGCCGTGCTGGTGGACCACGCCCTCAGCAGCGGCGACGCG 1891

Qy       1905 TACTGCCGTCAACGCGTGGAAGGTTGCCTCGGCCGAGAAGGCCATCTCGCTCACGCGCGA 1964
Db       1892 GAGCGGGAGCCCTCCGTGTTCTCCAAGATCACCAGGTTCGAGGAGGAGCTCCGCGCGGTG 1951

Qy       1965 GGTGCGCAACCGCTTCTGGCAGACGCCGTCTTCGCAGGCGCCGGCGCACGCATACCTCTC 2024
Db       1952 CTGCCCCAGGAGGTGGAGGCCGCCCGCGTGGCGGTCGCCGAGGGCAC-CGCCCCCGTGGC 2010
```

```
Qy       2025 GCCGCGCACGCGCGTCCTGTACTCGTTCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCG 2084
Db       2011 GAACCGGATCGCGGACAGCCGGTCGTTCCCGCTGTACCGCTTCGTGCGCGAGGAGCTCGG 2070
Qy       2085 CGGCGACGTGTTTGTCGGCGTGCAGCAGGAGACGATCGGGAGCAACGTCTCGCGCATCTA 2144
Db       2071 CTGCGTGTTCCTGACCGGCGAGAGGCTCAAGTCCCCCGGCGAGGAGTGCAACAAGGTGTT 2130
Qy       2145 CGAGGCCATCAAGGACGGCCGCATCAACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCC 2204
Db       2131 CGTCGGCATCAGCCAGGGCAAGCTCGTGGACCCCATGCTCGAGTGCCTCAAGGAGTGGGA 2190
Qy       2205 GAGCAAGCCTC 2215
Db       2191 CGGCAAGCCGC 2201
```

```
RESULT 2
AY106831
LOCUS       AY106831                2299 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO142079 mRNA sequence.
ACCESSION   AY106831
VERSION     AY106831.1  GI:21209909
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 2299)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 2299)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..2299
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638374"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match            16.0%;  Score 386.2;  DB 11;  Length 2299;
  Best Local Similarity  55.1%;  Pred. No. 5e-43;
  Matches  950;  Conservative    0;  Mismatches  753;  Indels   22;  Gaps    9;
```

```
Qy        503 TCAGCGTCGGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGG 562
Db        246 TCTTCGGCAACGGCTCCGACGGCCACACGCTCCCGTCCGAGGTTTCGCGTGCGGCCATGC 305
Qy        563 TCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGG 622
Db        306 TCGTCCGCATCAACACCCTCCTCCAGGGCTACTCCGGCATCCGCTTCGAGATCCTCGAGG 365
Qy        623 CGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCT 682
Db        366 CCATCACCAAGCTCATCAACACCGGCGTCA-GCCGTGCCTGCCGCTGCGGGGCACCATCA 424
Qy        683 CGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCG 742
Db        425 CCGCGTCGGGCGACCTCGTCCCGCTGTCCTACATCGCCGGCCTCATCACGGGGCGTCCCA 484
Qy        743 ACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCA 802
Db        485 ACGCGCAGGCCGTCACGGTTGACGGCAGGAAGGTGGACGCCGCCGAGGCGTTCAAGGCGG 544
Qy        803 TCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCA 862
Db        545 CCGGCATCGAGGG---CGGCTTCTTCAAGCTCAACCCCAAGGAGGGCCTGGCCATCGTCA 601
Qy        863 ACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCT 922
Db        602 ACGGCACGTCCGTGGGGTCGGCGCTCGCCGCCATGGTGTGCTTCGACGCCAACGTCCTCG 661
Qy        923 CGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCT 982
Db        662 CCGTCCTGTCCGAGGTCCTGTCCGCCGTCTTCTGCGAGGTGATGAACGGCAAGCCCGAGT 721
Qy        983 CGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGC 1042
Db        722 ACACGGACCACCTTACCCACAAGCTCAAGCAC---CACCCGGGGTCCATCGAGGCCGCGG 778
Qy       1043 GCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCA 1102
Db        779 CCATCATGGAGCACATCCTGGATGGCAGCTCCTT---CATGAAGCACGCCAAGGAGGTGA 835
Qy       1103 AGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGCTACCGCTCCGCACGTCGCCTC 1162
Db        836 ACGCCATGGACCCGCTGCTGAAGCCGAAGCAGGACAGGTACGCGCTCCGCACGTCGCCGC 895
Qy       1163 AGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGA 1222
Db        896 AGTGGCTCGGCCCCCAGATCGAGGTCATCCGCGCCGCCACCAAGTCCATCGAGCGAGAGG 955
Qy       1223 ACAACACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCG 1282
Db        956 TCAACTCGGTCAACGACAACCCTGTCATCGACGTCCACCGTGGCAAGGCGCTGCACGGCG 1015
Qy       1283 GCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCC 1342
Db       1016 GCAACTTCCAGGGCACCCCGATCGGCGTGTCCATGGACAACGCCCGCCTCGCCGTCGCCA 1075
Qy       1343 TCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCG 1402
Db       1076 ACATCGGCAAGCTCATGTTCGCCCAGTTCTCCGAGCTGGTCAACGAGTTCTACAACAACG 1135
Qy       1403 GCCTGCCTTCGTGCCTCGCTG---CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCT 1459
Db       1136 GGCTGACCTCCAACCTGGCGGGCAGCCGCAACCCGAGCCTGGACTACGGGTTCAAGGGCA 1195
Qy       1460 TGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTA 1519
Db       1196 CCGAGATCGCCATGGCGTCCTACTGCTCCGAGCTCCAGTACCTGGCCAACCCCATCACCA 1255
Qy       1520 CCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCG 1579
Db       1256 ACCACGTCCAGAGCGCGGAGCAGCACAACCAGGACGTGAACTCCCTCGGCCTCGTCTCCG 1315
Qy       1580 CGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACT 1639
Db       1316 CCAGGAAGACCGCCGAGGCGGTGGACATCCTCAAGCTCATGTCGTCCACGTACATGGTCG 1375
Qy       1640 GCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACC 1699
```

```
Db     1376 CGCTGTGCCAGGCCGTGGACCTGCGCCACCTCGAGGAGAACCTCAAGAGCGCCGTCAAGA 1435

Qy     1700 CGCTTCTCCCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTCAACGCAC--- 1756

Db     1436 GCTGCGTGATGGCGGTGGCCAGGAAGGTGCTGACCACCAGCCTCGGCGGCACCTCCACA 1495

Qy     1757 TTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACG--- 1813

Db     1496 GCGCGCGCTTCAGCGAGAAGGCCCTGCTCACCGCCATCGACCGCGAGGCGGTGTACGGGT 1555

Qy     1814 ACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCC 1873

Db     1556 ACTACGACGACCCCTGCAGCGCCAACTCGCCCCTGATGAAGAAGATCCGGGCCGTGCTGG 1615

Qy     1874 TCTCGTCCTCGCCCTCTGCCAACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933

Db     1616 TGGACCACGCCCTCGCCAGCGGCGAGGCCGAGAAGGACGCCAGCGCGTCCGTGTTCTCCA 1675

Qy     1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGC-GCAACCGCTTCTGGCAGACGCCG 1992

Db     1676 AGATCAACAGGTTCGAGGAGGAGCTGCGCGAGGCGCTGCCCAGGGAGATGGAGGCCGCCC 1735

Qy     1993 TCTTCGCAGGCGCCGGCGCACGCATACCTCTCGCC--GCGCACGCGCGTCCTGTACTCGT 2050

Db     1736 GCGTGGCCTTCGAGACCGGCGCCGCGCCGATCGCCAACAGGATCAAGGAGAGCCGGTCGT 1795

Qy     2051 TCGTGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGC 2110

Db     1796 ACCCGCTGTACCGCTTCATCCGCCAGGACCTCGGCGCCGTGTACCTGACCGGCGAGAAGC 1855

Qy     2111 AGGAGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCA 2170

Db     1856 TCAAGTCCCCCGGCGAGGAGTGCAACAAGGTGTTCCTGGCCCTCAGCGAGGGCAAGCTCA 1915

Qy     2171 ACCACGTCCTCGTCAAGATGCTCGCGTAAGGCCCGAGCAAGCCTC 2215

Db     1916 TCGACCCCATGCTCGACTGCCTCAAGGAGTGGGACGGCAAGCCCC 1960
```

```
RESULT 3
AY104679
LOCUS       AY104679                1864 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO142078 mRNA sequence.
ACCESSION   AY104679
VERSION     AY104679.1  GI:21207757
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1864)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1864)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1..1864
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:638373"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
ORIGIN

  Query Match             12.0%;  Score 289.2;  DB 11;  Length 1864;
  Best Local Similarity   57.8%;  Pred. No. 7.8e-30;
  Matches  594;  Conservative    0;  Mismatches  423;  Indels   11;  Gaps    4;

Qy      662 TCCCCCTCCGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCG 721

Db        1 TGCCGCTCCGGGGCACCATCACCGCGTCGGGCGACCTCGTCCCGCTCTCATACATCGCCG 60

Qy      722 GCGCCATCACCGGTCACCCCGACGTCAAGGTTCACGTTTGCACGAGGGAACCGAGAAGA 781

Db       61 GCCTCATCACGGGCCGCCAGAACTCC---GTCGCGGTGGGCCCCGACGGCAGGAAGGTGG 117

Qy      782 TCATGTTTGCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCGA 841

Db      118 GCGCCGCCGAGGCGTTCAGGATCGCCGGCATCGAGCACGGGTTCTTCGAGCTGCAGCCCA 177

Qy      842 AGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGTC 901

Db      178 AGGAGGGCCTCGCCATGGTCAACGGCACCGCCGTGGGCTCCGGCCTCGCCTCCACCGTGC 237

Qy      902 TGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAGG 961

Db      238 TCTTCGAGGCCAACGTTCTCGCCGTCATGGCCGAGGTCATCTCCGCGGTGTTCTGCGAGG 297

Qy      962 CCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCACC 1021

Db      298 TCATGACCGGCAAGCCCGAGTTCACCGACCACCTGACGCACAAGCTGAAGCAC---CACC 354

Qy     1022 CCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCCG 1081

Db      355 CCGGACAGATCGAGGCCGCCGCTATCATGGAGCACGTCCTGGAAGGCAGCTCGTACATGA 414

Qy     1082 TTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGCT 1141

Db      415 AGCTGGCCAAGAAGCTCGGCGAGCTTCGACCCGCTGATG--AAGCCGAAGCAGGACAGAT 472

Qy     1142 ACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCCT 1201

Db      473 ACGCGCTCCGCACCTCCCCGCAGTGGCTGGGCCCGCAGATCGAAGTGATCCGCTTTGCCA 532

Qy     1202 ACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAGA 1261

Db      533 CCAAGTCCATCGAGCGCGAGATCAACTCCGTCAACGACAACCCGCTCATCGATGTGTCCC 592

Qy     1262 ACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGA 1321

Db      593 GTGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGAACTCCCATCGGCGTGTCCATGGACA 652

Qy     1322 AGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTGC 1381

Db      653 ACACCCGTCTCGCGCTCGCCGCCATCGGCAAGCTCATGTTCGCGCAGTTCTCGGAGCTGG 712

Qy     1382 TCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTGCCGAG---GACCCGTCGC 1438

Db      713 TGAACGACTACTACAACAACGGCCTCCCGTCCAACCTGTCCGGCGGGCGCAACCCCAGCT 772

Qy     1439 TCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCGGCC 1498

Db      773 TGGACTACGGCTTCAAGGGCGCCGAGATCGCCATGGCCTCCTACTGCTCGGAGCTCCAGT 832

Qy     1499 ACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCGTCA 1558
```

```
Db         833 TCCTGGGCAACCCGGTGACCAACCACGTCCAGAGCGCGGAGCAGCACAACCAGGACGTGA 892

Qy        1559 ACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTCTCC 1618

Db         893 ACTCCCTGGGCCTCATCTCGTCCAGGAAGACCGCCGAGGCAGTGGACATCCTGAAGCTGA 952

Qy        1619 TTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAGCTCG 1678

Db         953 TGACGTCCACGTTCCTGATCGCGCTGTGCCAGGCCATCGACCTGCGGCACCTCGAGGAGA 1012

Qy        1679 ACTTCAAG 1686

Db        1013 ACGTGAAG 1020


RESULT 4
CB632830
LOCUS       CB632830                 859 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSIIEb11F05.f OSIIEb Oryza sativa (indica cultivar-group) cDNA
            clone OSIIEb11F05 5', mRNA sequence.
ACCESSION   CB632830
VERSION     CB632830.1  GI:29627819
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 859)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 11  row: F  column: 05
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR36"
                     /db_xref="taxon:39946"
                     /clone="OSIIEb11F05"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSIIEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (PO6-6-3)"
ORIGIN

  Query Match             11.1%;  Score 268;  DB 14;  Length 859;
  Best Local Similarity   59.2%;  Pred. No. 6e-27;
  Matches  514;  Conservative    0;  Mismatches  345;  Indels    9;  Gaps    3;

Qy         550 CGCGGCGCCATGGTCATCCGCGTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTC 609

Db           1 CGGCGGCGCCATGCTCGTGCGCATCAACACCCTCCTCCAGGGCTACTCCGGCATCCGGTTC 60

Qy         610 GTCGTCCTTGAGGCGCTCACCAACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTC 669

Db          61 GAGATCCTCGAGGCCATCACCAAGCTGCTCAACACCGGCGTCACGCCGTGCCTGCCGCTC 120

Qy         670 CGCGGCTCCATCTCGGCGTCGGGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATC 729

Db         121 CGTGGCACCATCACCGCGTCCGGTGACCTGGTTCCCCTGTCCTACATTGCCGGCCTCATC 180

Qy         730 ACCGGTCACCCCGACGTCAAGGTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTT 789

Db         181 ACCGGCCGCCCCAACGCGCAGG---CCATCTCGCCCGACGGCAGGAAGGTGGACGCCGCC 237

Qy         790 GCGCGCGAGGCCATCTCGCTCTTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGT 849

Db         238 GAGGCGTTCAAGCTCGCCGGCATCGAGGGTGGCTTCTTCACGCTGAACCCCAAGGAAGGT 297

Qy         850 CTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGAC 909

Db         298 CTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCGGCCACCGTGATGTTCGAC 357

Qy         910 TCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTC 969

Db         358 GCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTGTTCTGCGAGGTGATGAAC 417

Qy         970 GGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAG 1029

Db         418 GGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAGCAC---CACCCTGGGTCG 474

Qy        1030 GTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCAC 1089

Db         475 ATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGCTCGTT---CATGAGCCAC 531

Qy        1090 GAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTC 1149

Db         532 GCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAGCAGGACAGGTACGCGCTC 591

Qy        1150 CGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACT 1209

Db         592 CGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATCCGCGCCGCCACCAAGTCC 651

Qy        1210 CTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAG 1269

Db         652 ATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATCGACGTCCACCGCGGCAAG 711

Qy        1270 ACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGG 1329

Db         712 GCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTGTCCATGGACAACGCCCGT 771

Qy        1330 CTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCT 1389

Db         772 CTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGAGCTCGTGAACGAA 831

Qy        1390 GCCATGAACCGCGGCCTGCCTTCGTGCC 1417

Db         832 AGTCTACACAACGGGCTGACCTCCAACC 859


RESULT 5
CB645202/c
LOCUS       CB645202                 866 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb06M01.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb06M01 3', mRNA sequence.
ACCESSION   CB645202
VERSION     CB645202.1  GI:29640193
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 866)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
```

```
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 06  row: M  column: 01
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1. .866
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb06M01"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             11.0%;  Score 266.8;  DB 14;  Length 866;
  Best Local Similarity   59.7%;  Pred. No. 8.7e-27;
  Matches  505;  Conservative    0;  Mismatches  332;  Indels    9;  Gaps    3;

Qy        832 CTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCG 891
Db        865 CTGAACCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCG 806

Qy        892 ACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTC 951
Db        805 GCCACCGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTG 746

Qy        952 ACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGC 1011
Db        745 TTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAG 686

Qy       1012 CGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCG 1071
Db        685 CAC---CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGC 629

Qy       1072 TCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGC 1131
Db        628 TCGTT---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAG 572

Qy       1132 CAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATG 1191
Db        571 CAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATC 512

Qy       1192 ATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTC 1251
Db        511 CGCGCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATC 452

Qy       1252 GACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATT 1311
Db        451 GACGTCCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTG 392

Qy       1312 TCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGC 1371
Db        391 TCCATGGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTC 332

Qy       1372 ACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAG 1428
Db        331 TCCGAGCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGC 272

Qy       1429 GACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCG 1488
Db        271 AACCCGAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCT 212

Qy       1489 GAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAAC 1548
Db        211 GAGCTCCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAAC 152

Qy       1549 CAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTC 1608
Db        151 CAGGACGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATC 92

Qy       1609 CTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCG 1668
Db         91 CTCAAGCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTCGACCTTCGCCAC 32

Qy       1669 ATGGAG 1674
Db         31 CTCGAG 26


RESULT 6
CB676595/c
LOCUS       CB676595                 860 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEe12O20.r OSJNEe Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEe12O20 3', mRNA sequence.
ACCESSION   CB676595
VERSION     CB676595.1  GI:29680320
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 860)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 12  row: O  column: 20
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1. .860
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEe12O20"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEe"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (70-15)"
ORIGIN

  Query Match             11.0%;  Score 266.6;  DB 14;  Length 860;
  Best Local Similarity   59.8%;  Pred. No. 9.2e-27;
  Matches  503;  Conservative    0;  Mismatches  329;  Indels    9;  Gaps    3;
```

```
Qy    837 CCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCT 896
Db    860 CCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCGGCCAC 801

Qy    897 CAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGT 956
Db    800 CGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTGTTCTG 741

Qy    957 GGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCC 1016
Db    740 CGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAGCAC-- 683

Qy   1017 GCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTT 1076
Db    682 -CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGCTCGTT 624

Qy   1077 TGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGA 1136
Db    623 ---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAGCAGGA 567

Qy   1137 CCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCA 1196
Db    566 CAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATCCGCGC 507

Qy   1197 CGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGT 1256
Db    506 CGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATCGACGT 447

Qy   1257 CGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGAT 1316
Db    446 CCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTGTCCAT 387

Qy   1317 GGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGA 1376
Db    386 GGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGA 327

Qy   1377 GTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAGGACCC 1433
Db    326 GCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGCAACCC 267

Qy   1434 GTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCT 1493
Db    266 GAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCTGAGCT 207

Qy   1494 CGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGC 1553
Db    206 CCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAACCAGGA 147

Qy   1554 CGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTC 1613
Db    146 CGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATCCTCAA 87

Qy   1614 TCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGA 1673
Db     86 GCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTCGACCTTCGCCACCTCGA 27

Qy   1674 G 1674
Db     26 G 26
```

RESULT 7
CB648876/c

```
LOCUS       CB648876                 889 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb12H22.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb12H22 3', mRNA sequence.
ACCESSION   CB648876
VERSION     CB648876.1  GI:29643869
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 12  row: H  column: 22
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..889
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb12H22"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             11.0%;  Score 265.6;  DB 14;  Length 889;
  Best Local Similarity   59.8%;  Pred. No. 1.3e-26;
  Matches  502;  Conservative    0;  Mismatches  329;  Indels    9;  Gaps    3;

Qy    838 CCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTC 897
Db    859 CCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCGGCCACC 800

Qy    898 AGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTG 957
Db    799 GTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTGTTCTGC 740

Qy    958 GAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCG 1017
Db    739 GAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAGCAC--- 683

Qy   1018 CACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTT 1077
Db    682 CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGCTCGTT- 624

Qy   1078 GCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGAC 1137
Db    623 --CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAGCAGGAC 566

Qy   1138 CGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCAC 1197
Db    565 AGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATCCGCGCC 506

Qy   1198 GCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTC 1257
Db    505 GCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATCGACGTC 446

Qy   1258 GAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATG 1317
Db    445 CACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTGTCCATG 386

Qy   1318 GAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAG 1377
```

```
Db        385 GACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGAG 326

Qy       1378 TTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAGGACCCG 1434

Db        325 CTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGCAACCCG 266

Qy       1435 TCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTC 1494

Db        265 AGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCTGAGCTC 206

Qy       1495 GGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCC 1554

Db        205 CAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAACCAGGAC 146

Qy       1555 GTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCT 1614

Db        145 GTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATCCTCAAG 86

Qy       1615 CTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAG 1674

Db         85 CTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTCGACCTTCGCCACCTCGAG 26


RESULT 8
CB649712/c
LOCUS       CB649712                 864 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb13M21.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb13M21 3', mRNA sequence.
ACCESSION   CB649712
VERSION     CB649712.1  GI:29644705
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 13  row: M  column: 21
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..864
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb13M21"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             11.0%;  Score 265.4;  DB 14;  Length 864;
  Best Local Similarity   59.7%;  Pred. No. 1.3e-26;
  Matches  503;  Conservative    0;  Mismatches  331;  Indels    9;  Gaps    3;

Qy        835 GGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACC 894

Db        862 GACCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCGGCC 803

Qy        895 CTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACG 954

Db        802 ACCGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTGTTC 743

Qy        955 GTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGC 1014

Db        742 TGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAGCAC 683

Qy       1015 CCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCG 1074

Db        682 ---CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGCTCG 626

Qy       1075 TTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAG 1134

Db        625 TT---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAGCAG 569

Qy       1135 GACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATG 1194

Db        568 GACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATCCGC 509

Qy       1195 CACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGAC 1254

Db        508 GCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATCGAC 449

Qy       1255 GTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCG 1314

Db        448 GTCCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTGTCC 389

Qy       1315 ATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACC 1374

Db        388 ATGGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCC 329

Qy       1375 GAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAGGAC 1431

Db        328 GAGCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGCAAC 269

Qy       1432 CCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAG 1491

Db        268 CCGAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCTGAG 209

Qy       1492 CTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAG 1551

Db        208 CTCCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAACCAG 149

Qy       1552 GCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTT 1611

Db        148 GACGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATCCTC 89

Qy       1612 TCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATG 1671

Db         88 AAGCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTAGACCTTCGCCACCTC 29

Qy       1672 GAG 1674
              |||
Db         28 GAG 26


RESULT 9
CB643557/c
LOCUS       CB643557                 867 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb04F02.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb04F02 3', mRNA sequence.
ACCESSION   CB643557
VERSION     CB643557.1  GI:29638548
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
```

```
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 04  row: F  column: 02
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb04F02"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             11.0%;  Score 265.2;  DB 14;  Length 867;
  Best Local Similarity   59.6%;  Pred. No. 1.4e-26;
  Matches  504;  Conservative    0;  Mismatches  333;  Indels    9;  Gaps    3;

Qy        832 CTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCG 891

Db        865 CTGAACCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCG 806

Qy        892 ACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTC 951

Db        805 GCCACCGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTG 746

Qy        952 ACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGC 1011

Db        745 TTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAG 686

Qy       1012 CGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCG 1071

Db        685 CAC---CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGC 629

Qy       1072 TCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGC 1131

Db        628 TCGTT---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAG 572

Qy       1132 CAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATG 1191

Db        571 CAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATC 512

Qy       1192 ATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTC 1251

Db        511 CGCGCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATC 452

Qy       1252 GACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATT 1311

Db        451 GACGTCCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTG 392

Qy       1312 TCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGC 1371

Db        391 TCCATGGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTC 332

Qy       1372 ACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAG 1428

Db        331 TCCGAGCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGC 272

Qy       1429 GACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCG 1488

Db        271 AACCCGAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCT 212

Qy       1489 GAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAAC 1548

Db        211 GAGCTCCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAAC 152

Qy       1549 CAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTC 1608

Db        151 CAGGACGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATC 92

Qy       1609 CTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCG 1668

Db         91 CTCAAGCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTCGACCGTCGCCAC 32

Qy       1669 ATGGAG 1674

Db         31 CTCGAG 26


RESULT 10
CB651503/c
LOCUS       CB651503                 867 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb16H19.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb16H19 3', mRNA sequence.
ACCESSION   CB651503
VERSION     CB651503.1  GI:29646496
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 16  row: H  column: 19
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb16H19"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
```

```
                    /clone_lib="OSJNEb"
                    /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                    XhoI; 24 hrs after innoculation with Rice Blast (Che
                    86061)"
ORIGIN

  Query Match             11.0%;  Score 265.2;  DB 14;  Length 867;
  Best Local Similarity   59.6%;  Pred. No. 1.4e-26;
  Matches  504;  Conservative    0;  Mismatches  333;  Indels    9;  Gaps    3;

Qy        832 CTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCG 891
Db        865 CTGAACCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCG 806

Qy        892 ACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTC 951
Db        805 GCCACCGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTG 746

Qy        952 ACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGC 1011
Db        745 TTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAG 686

Qy       1012 CGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCG 1071
Db        685 CAC---CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGC 629

Qy       1072 TCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGC 1131
Db        628 TCGTT---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAG 572

Qy       1132 CAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATG 1191
Db        571 CAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATC 512

Qy       1192 ATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTC 1251
Db        511 CGCGCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATC 452

Qy       1252 GACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATT 1311
Db        451 GACGTCCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTG 392

Qy       1312 TCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGC 1371
Db        391 TCCATGGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTC 332

Qy       1372 ACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAG 1428
Db        331 TCCGAGCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGC 272

Qy       1429 GACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCG 1488
Db        271 AACCCGAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCT 212

Qy       1489 GAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAAC 1548
Db        211 GAGCTCCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAAC 152

Qy       1549 CAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTC 1608
Db        151 CAGGACGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATC 92

Qy       1609 CTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCG 1668
Db         91 CTCAAGCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGACGACCTTCGCCAC 32

Qy       1669 ATGGAG 1674
Db         31 CTCGAG 26

RESULT 11
CB658123/c
LOCUS       CB658123                 860 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEc13022.r OSJNEc Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEc13022 3', mRNA sequence.
ACCESSION   CB658123
VERSION     CB658123.1  GI:29661848
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 860)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 13  row: O  column: 22
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..860
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEc13022"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEc"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 6 hrs after innoculation with Rice Blast (C9240-1)"
ORIGIN

  Query Match             11.0%;  Score 265;  DB 14;  Length 860;
  Best Local Similarity   59.7%;  Pred. No. 1.5e-26;
  Matches  502;  Conservative    0;  Mismatches  330;  Indels    9;  Gaps    3;

Qy        837 CCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCT 896
Db        860 CCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCGGCCAC 801

Qy        897 CAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGT 956
Db        800 CGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTGTTCTG 741

Qy        957 GGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCC 1016
Db        740 CGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAGCAC-- 683

Qy       1017 GCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTT 1076
Db        682 -CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGCTCGTT 624

Qy       1077 TGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGA 1136
Db        623 ---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAGCAGGA 567

Qy       1137 CCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCA 1196
Db        566 CAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATCCGCGC 507
```

```
Qy       1197 CGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGT 1256

Db        506 CGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATCGACGT 447

Qy       1257 CGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGAT 1316

Db        446 CCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTGTCCAT 387

Qy       1317 GGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGA 1376

Db        386 GGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGA 327

Qy       1377 GTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAGGACCC 1433

Db        326 GCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGCAACCC 267

Qy       1434 GTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCT 1493

Db        266 GAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCTGAGCT 207

Qy       1494 CGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGC 1553

Db        206 CCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAACCAGGA 147

Qy       1554 CGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTC 1613

Db        146 CGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATCCTCAA 87

Qy       1614 TCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGA 1673

Db         86 GCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTCGAGCTTCGCCACCTCGA 27

Qy       1674 G 1674

Db         26 G 26
```

```
RESULT 12
CB651685/c
LOCUS       CB651685                 891 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb16L23.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb16L23 3', mRNA sequence.
ACCESSION   CB651685
VERSION     CB651685.1  GI:29646678
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 891)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 16  row: L  column: 23
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..891
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb16L23"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             10.9%;  Score 263.8;  DB 14;  Length 891;
  Best Local Similarity   59.3%;  Pred. No. 2.2e-26;
  Matches  505;  Conservative    0;  Mismatches  337;  Indels    9;  Gaps    3;
```

```
Qy        827 TCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGA 886

Db        870 TCACGCTGAACCCCAAGAAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGC 811

Qy        887 TGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGG 946

Db        810 TCGCGGCCACCGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGG 751

Qy        947 CTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACG 1006

Db        750 CGGTGTTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGC 691

Qy       1007 TCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCG 1066

Db        690 TGAAGCAC---CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCG 634

Qy       1067 GCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTC 1126

Db        633 GGAGCTCGTT---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGC 577

Qy       1127 TTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGG 1186

Db        576 CGAAGCAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGG 517

Qy       1187 ACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGC 1246

Db        516 TCATCCGCGCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGG 457

Qy       1247 TCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCT 1306

Db        456 TGATCGACGTCCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCG 397

Qy       1307 CGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGC 1366

Db        396 GTGTGTCCATGGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGC 337

Qy       1367 AGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG--- 1423

Db        336 AGTTCTCCGAGCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCA 277

Qy       1424 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1483

Db        276 GCCGCAACCCGAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACT 217

Qy       1484 CTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGG 1543

Db        216 GCTCTGAGCTCCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGC 157

Qy       1544 GCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACG 1603

Db        156 ACAACCAGGACGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGG 97

Qy       1604 ACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCC 1663

Db         96 ACATCCTCAAGCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGACGACCTTC 37

Qy       1664 GCGCGATGGAG 1674
```

```
Db         36 GCCACCTCGAG 26

RESULT 13
CB648470/c
LOCUS       CB648470                 893 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb11N20.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb11N20 3', mRNA sequence.
ACCESSION   CB648470
VERSION     CB648470.1  GI:29643463
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 11  row: N  column: 20
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..893
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb11N20"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             10.9%;  Score 263.8;  DB 14;  Length 893;
  Best Local Similarity   59.3%;  Pred. No. 2.2e-26;
  Matches  505;  Conservative    0;  Mismatches  337;  Indels    9;  Gaps    3;

Qy        827 TCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGA 886
Db        870 TCACGCTGAACCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGGGGGGTCGGCGC 811

Qy        887 TGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGG 946
Db        810 TCGCGGCCACCGTGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGG 751

Qy        947 CTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACG 1006
Db        750 CGGTGTTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGC 691

Qy       1007 TCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCG 1066
Db        690 TGAAGCAC---CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCG 634

Qy       1067 GCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTC 1126
Db        633 GGAGCTCGTT---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGC 577

Qy       1127 TTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGG 1186
Db        576 CGAAGCAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGG 517

Qy       1187 ACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGC 1246
Db        516 TCATCCGCGCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGG 457

Qy       1247 TCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCT 1306
Db        456 TGATCGACGTCCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCG 397

Qy       1307 CGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGC 1366
Db        396 GTGTGTCCATGGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGC 337

Qy       1367 AGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG--- 1423
Db        336 AGTTCTCCGAGCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCA 277

Qy       1424 CCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACG 1483
Db        276 GCCGCAACCCGAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACT 217

Qy       1484 CTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGG 1543
Db        216 GCTCTGAGCTCCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGC 157

Qy       1544 GCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACG 1603
Db        156 ACAACCAGGACGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGG 97

Qy       1604 ACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCC 1663
Db         96 ACATCCTCAAGCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTAGACCTTC 37

Qy       1664 GCGCGATGGAG 1674
Db         36 GCCACCTCGAG 26

RESULT 14
CB673619/c
LOCUS       CB673619                 868 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEe08D18.r OSJNEe Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEe08D18 3', mRNA sequence.
ACCESSION   CB673619
VERSION     CB673619.1  GI:29677344
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 868)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
```

```
            BACKWARD: gga aac agc tat gac cat g
            Plate: 08  row: D  column: 18
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..868
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEe08D18"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEe"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (70-15)"
ORIGIN

  Query Match             10.9%;  Score 263.6;  DB 14;  Length 868;
  Best Local Similarity   59.5%;  Pred. No. 2.4e-26;
  Matches  503;  Conservative    0;  Mismatches  334;  Indels    9;  Gaps    3;

Qy        832 CTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCG 891
Db        865 CTGAACCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCGTGGGGTCGGCGCTCGCG 806

Qy        892 ACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTC 951
Db        805 GCCACCGTGATGTTTGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTG 746

Qy        952 ACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGC 1011
Db        745 TTCTGCGAGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAG 686

Qy       1012 CGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCG 1071
Db        685 CAC---CACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGC 629

Qy       1072 TCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGC 1131
Db        628 TCGTT---CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAG 572

Qy       1132 CAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATG 1191
Db        571 CAGGACAGGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATC 512

Qy       1192 ATGCACGCCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTC 1251
Db        511 CGCGCCGCCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATC 452

Qy       1252 GACGTCGAGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATT 1311
Db        451 GACGTCCACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTG 392

Qy       1312 TCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGC 1371
Db        391 TCCATGGACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTC 332

Qy       1372 ACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAG 1428
Db        331 TCCGAGCTCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGC 272

Qy       1429 GACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCG 1488
Db        271 AACCCGAGCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCT 212

Qy       1489 GAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAAC 1548
Db        211 GAGCTCCAGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAAC 152

Qy       1549 CAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTC 1608
Db        151 CAGGACGTGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATC 92

Qy       1609 CTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCG 1668
Db         91 CTCAAGCTCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTAGACCTTCGCCAC 32

Qy       1669 ATGGAG 1674
Db         31 CTCGAG 26


RESULT 15
CB647704/c
LOCUS       CB647704                 858 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEb10L21.r OSJNEb Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEb10L21 3', mRNA sequence.
ACCESSION   CB647704
VERSION     CB647704.1  GI:29642697
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 10  row: L  column: 21
            Seq primer: gga aac agc tat gac cat g.
FEATURES             Location/Qualifiers
     source          1..858
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEb10L21"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEb"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

  Query Match             10.9%;  Score 263;  DB 14;  Length 858;
  Best Local Similarity   59.6%;  Pred. No. 2.8e-26;
  Matches  500;  Conservative    0;  Mismatches  330;  Indels    9;  Gaps    3;

Qy        839 CGAAGGAGGGTCTCGGTCTGGTCAACGGAACGGCCGTCTCCGCCTCGATGGCGACCCTCA 898
Db        858 CCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCTGTGGGGTCGGCGCTCGCGGCCACCG 799

Qy        899 GTCTGCACGACTCGCACATGCTCTCGCTCCTCTCGCAGGCCTTGACGGCTCTCACGGTGG 958
Db        798 TGATGTTCGACGCCAACATCCTCGCCGTCCTGTCCGAGGTGCTCTCGGCGGTGTTCTGCG 739

Qy        959 AGGCCATGGTCGGCCAGCAGGGCTCGTTCGCGCCGTTCATCCACGACGTCTGCCGCCCGC 1018
Db        738 AGGTGATGAACGGCAAGCCGGAGTACACCGACCACCTGACCCACAAGCTGAAGCAC---C 682
```

```
Qy       1019 ACCCCGGCCAGGTCGAGGTCGCGCGCAACATCCGCACGCTCCTTTCCGGCTCGTCGTTTG 1078

Db        681 ACCCTGGGTCGATCGAGGCCGCCGCCATCATGGAGCACATCCTCGCCGGGAGCTCGTT-- 624

Qy       1079 CCGTTGAGCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTTCGCCAGGACC 1138

Db        623 -CATGAGCCACGCCAAGAAGGTGAACGAGATGGACCCGCTGCTGAAGCCGAAGCAGGACA 565

Qy       1139 GCTACCCGCTCCGCACGTCGCCTCAGTTCCTCGGCCCGCTCGTGGAGGACATGATGCACG 1198

Db        564 GGTACGCGCTCCGCACGTCGCCGCAGTGGCTCGGCCCGCAGATCGAGGTCATCCGCGCCG 505

Qy       1199 CCTACTCGACTCTCTCGCTCGAGAACAACACGACGACCGACAACCCGCTCCTCGACGTCG 1258

Db        504 CCACCAAGTCCATCGAGCGCGAGGTCAACTCCGTGAACGACAACCCGGTGATCGACGTCC 445

Qy       1259 AGAACAAGCAGACCGCGCACGGCGGCAACTTCCAGGCGTCGGCTGTCTCGATTTCGATGG 1318

Db        444 ACCGCGGCAAGGCGCTCCACGGCGGCAACTTCCAGGGCACCCCCATCGGTGTGTCCATGG 385

Qy       1319 AGAAGACCAGGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCACGCAGTGCACCGAGT 1378

Db        384 ACAACGCCCGTCTCGCCATCGCCAACATCGGCAAGCTCATGTTCGCGCAGTTCTCCGAGC 325

Qy       1379 TGCTCAACGCTGCCATGAACCGCGGCCTGCCTTCGTGCCTCGCTG---CCGAGGACCCGT 1435

Db        324 TCGTGAACGAGTTCTACAACAACGGGCTGACCTCCAACCTGGCCGGCAGCCGCAACCCGA 265

Qy       1436 CGCTCAACTATCACGGCAAGGGCTTGGACATTCACATCGCTGCTTACGCTTCGGAGCTCG 1495

Db        264 GCTTGGACTACGGGTTCAAGGGCACCGAGATCGCCATGGCCTCCTACTGCTCTGAGCTCC 205

Qy       1496 GCCACCTTGCCAACCCGGTCACTACCTTCGTCCAGCCCGCAGAGATGGGCAACCAGGCCG 1555

Db        204 AGTACCTCGCCAACCCCATCACCAACCATGTCCAGAGCGCGGAGCAGCACAACCAGGACG 145

Qy       1556 TCAACTCGCTCGCTCTCATCTCCGCGCGCCGCACTGCCGAGGCCAACGACGTCCTTTCTC 1615

Db        144 TGAACTCGCTGGGTCTCGTCTCGGCCAGGAAGACCCTGGAGGCGGTGGACATCCTCAAGC 85

Qy       1616 TCCTTCTCGCCTCGCACCTGTACTGCACGCTCCAGGCCGTCGACCTCCGCGCGATGGAG 1674

Db         84 TCATGACCTCCACCTACATCGTCGCCCTGTGCCAGGCCGTCGACCTTCGCCACCTCGAG 26
```

Search completed: September 12, 2004, 04:20:22
Job time : 9130.4 secs

```
OM protein - protein search, using sw model

Run on:         September  9, 2004, 09:49:43 ; Search time 88.6307 Seconds
                                          (without alignments)
                                          2295.302 Million cell updates/sec

Title:          US-09-939-408A-13
Perfect score:  3610
Sequence:       1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3608	99.9	720	5	ABB07690	Abb07690 R. gramin
2	3608	99.9	720	5	AAE27938	Aae27938 Rhodotoru
3	2770	76.7	726	5	ABB07694	Abb07694 Rhodotoru
4	2770	76.7	726	5	AAE27942	Aae27942 Phenylala
5	2730	75.6	720	5	AAE27943	Aae27943 Yeast phe
6	2619	72.5	716	1	AAP83141	Aap83141 Sequence
7	2614	72.4	716	1	AAP80513	Aap80513 L-phenyla
8	2614	72.4	716	5	ABB07693	Abb07693 R. torulo
9	2614	72.4	716	5	AAE27941	Aae27941 Rhodotoru
10	2614	72.4	716	5	AAE20653	Aae20653 Rhodotoru
11	2614	72.4	716	5	AAE16389	Aae16389 Rhodotoru
12	2614	72.4	716	6	ABG71948	Abg71948 R. glutin
13	2614	72.4	716	7	ADE36054	Ade36054 Rhodospor
14	2613	72.4	716	5	AAE20660	Aae20660 Rhodotoru
15	2613	72.4	716	6	ABG71951	Abg71951 R. glutin
16	2612	72.4	716	5	AAE20665	Aae20665 R. glutin
17	2612	72.4	716	5	AAE20668	Aae20668 R. glutin
18	2612	72.4	716	5	AAE20654	Aae20654 Rhodotoru
19	2612	72.4	716	5	AAE16390	Aae16390 Rhodotoru
20	2612	72.4	716	6	ABG71955	Abg71955 R. glutin
21	2612	72.4	716	7	ADE36056	Ade36056 Rhodospor
22	2611	72.3	716	1	AAP81099	Aap81099 Sequence
23	2609	72.3	716	7	ADE36076	Ade36076 Rhodospor
24	2608	72.2	716	7	ADE36074	Ade36074 Rhodospor
25	2607	72.2	716	5	AAE20663	Aae20663 Rhodotoru
26	2607	72.2	716	6	ABG71954	Abg71954 R. glutin
27	2607	72.2	716	7	ADE36068	Ade36068 Rhodospor
28	2606	72.2	716	5	AAE20667	Aae20667 R. glutin
29	2606	72.2	716	7	ADE36060	Ade36060 Rhodospor
30	2605	72.2	716	5	AAE20711	Aae20711 R. glutin
31	2605	72.2	716	7	ADE36062	Ade36062 Rhodospor
32	2604	72.1	716	7	ADE36066	Ade36066 Rhodospor
33	2603	72.1	716	5	AAE20662	Aae20662 Rhodotoru
34	2603	72.1	716	5	AAE20661	Aae20661 Rhodotoru
35	2603	72.1	716	5	AAE20666	Aae20666 R. glutin
36	2603	72.1	716	6	ABG71952	Abg71952 R. glutin
37	2603	72.1	716	6	ABG71953	Abg71953 R. glutin
38	2601	72.0	716	7	ADE36064	Ade36064 Rhodospor
39	2600	72.0	716	1	AAP95783	Aap95783 L-phenyla
40	2599	72.0	716	5	AAE20659	Aae20659 Rhodotoru
41	2599	72.0	716	6	ABG71950	Abg71950 R. glutin
42	2599	72.0	716	7	ADE36070	Ade36070 Rhodospor
43	2598	72.0	716	5	AAE20664	Aae20664 R. glutin
44	2598	72.0	716	7	ADE36058	Ade36058 Rhodospor
45	2598	72.0	716	7	ADE36072	Ade36072 Rhodospor

ALIGNMENTS

```
RESULT 1
ABB07690
ID   ABB07690 standard; protein; 720 AA.
XX
AC   ABB07690;
XX
DT   10-JUN-2002  (first entry)
XX
DE   R. graminis PAL polypeptide.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; enzyme;
KW   gene therapy.
XX
OS   Rhodotorula graminis.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  153
FT                    /label= Val or Ala
FT                    /note= "encoded by GYC"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   N-PSDB; ABA95240, ABA95251.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Claim 9; Fig 1; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
```

```
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a R. graminis PAL polypeptide
XX
SQ   Sequence 720 AA;

  Query Match             99.9%;  Score 3608;  DB 5;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-306;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720


RESULT 2
AAE27938
ID   AAE27938 standard; protein; 720 AA.
XX
AC   AAE27938;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Rhodotorula graminis PAL protein #1.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 153
FT                   /label= Val, Ala
FT                   /note= "Encoded by GYC"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   N-PSDB; AAD45812.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 26; Page 29-31; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is R. graminis PAL
CC   protein
XX
SQ   Sequence 720 AA;

  Query Match             99.9%;  Score 3608;  DB 5;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-306;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
```

```
RESULT 3
ABB07694
ID   ABB07694 standard; protein; 726 AA.
XX
AC   ABB07694;
XX
DT   29-AUG-2003  (revised)
DT   07-AUG-2003  (revised)
DT   10-JUN-2002  (first entry)
XX
DE   Rhodotorula PAL consensus polypeptide sequence.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
KW   consensus; enzyme.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .726
FT                   /note= "Xaa is any amino acid; "Xaa" indicates no
FT                   consensus at that position"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   N-PSDB; ABA95244.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Claim 10; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a consensus polypeptide sequence of the Rhodotorula PAL.
CC   (Updated on 07-AUG-2003 to correct OS field.) (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 726 AA;

  Query Match             76.7%;  Score 2770;  DB 5;  Length 726;
  Best Local Similarity   77.7%;  Pred. No. 3.7e-233;
  Matches  564;  Conservative   40;  Mismatches  116;  Indels    6;  Gaps    3;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:||: |||  || |||    |           :||:||:  |:|
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy         61 P-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVT 119
              | |||  || |||||: |||||||||| |||  :  ||||  ::||||:||: || ||||||
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        120 TGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRV 179
              ||||||||||||:||||||:||||||| |||| ||  :||||||:||||||||| |||
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239
              |||||||||||||||||||||| ||||||||:|||||||||||||| |||||| ||
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        240 HVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
              || ||| |||  |||||:|||| ||||||||||||||||||||||||||:|||:||||||
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
              ||||||||||||  ||| ||:||| |||| |:|||||||||| || ||| ||||||||
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNF 418
              |||||||||||||||:||||| ||:||:: ||||    :|||||:||||| | |||||
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIH 478
              ||||  :||||||||||||||||||| ||:||| |||||||||||||||||:|| |||||
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
               ||| |||||||||||| ||||||||||||||||||||||||||||||||||:||||  ||
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA----LALEVKKALNKRLEQTTTYDL 594
              |||||||  :||||| |     || || |           |  :| |  | |||||| :|||
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy        595 EPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSS 654
              |||||||| |||||||| |||           |:| |||||||| ||||||||| ||  ||  |||
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        655 QAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHV 714
                :||  ||||||||||||||||||:||||||:| |: |||||||||||||| |||| |
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        715 LVKMLA 720
              ||||||
Db        721 LVKMLA 726

RESULT 4
```

```
AAE27942
ID   AAE27942 standard; protein; 726 AA.
XX
AC   AAE27942;
XX
DT   06-AUG-2003  (revised)
DT   27-DEC-2002  (first entry)
XX
DE   Phenylalanine ammonia lyase consensus protein.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 726. .719
FT                   /note= "Xaa corresponds to unknown amino acid residue"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 26; Page 47-48; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is PAL consensus protein.
CC   (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 726 AA;

  Query Match             76.7%;  Score 2770;  DB 5;  Length 726;
  Best Local Similarity   77.7%;  Pred. No. 3.7e-233;
  Matches  564;  Conservative   40;  Mismatches  116;  Indels    6;  Gaps    3;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy         61 P-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVT 119
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        120 TGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRV 179
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        240 HVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNF 418
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIH 478
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA----LALEVKKALNKRLEQTTTYDL 594
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXXXLXKRLEQTNSYDL 600

Qy        595 EPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSS 654
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        655 QAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHV 714
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        715 LVKMLA 720
Db        721 LVKMLA 726


RESULT 5
AAE27943
ID   AAE27943 standard; protein; 720 AA.
XX
AC   AAE27943;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Yeast phenylalanine ammonia lyase protein #2.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 5
FT                   /label= Val, Leu, Phe
FT                   /note= "Encoded by BTC"
FT   Misc-difference 12
FT                   /label= Val, Leu, Phe
FT                   /note= "Encoded by BTC"
FT   Misc-difference 16
FT                   /label= Val, Leu, Phe
FT                   /note= "Encoded by BTC"
FT   Misc-difference 17
FT                   /label= Thr, Ala, Ser
FT                   /note= "Encoded by DCV"
FT   Misc-difference 19
FT                   /label= Gly
FT                   /note= "Encoded by GGV"
FT   Misc-difference 20
FT                   /label= Ser, Leu
```

```
FT                     /note= "Encoded by THG or an in-frame stop codon"
FT   Misc-difference 25
FT                     /label= Thr, Pro, Ser
FT                     /note= "Encoded by HCC"
FT   Misc-difference 27
FT                     /label= Pro, Ser
FT                     /note= "Encoded by YCG"
FT   Misc-difference 28
FT                     /label= Ala, Pro
FT                     /note= "Encoded by SCW"
FT   Misc-difference 34
FT                     /label= Thr
FT                     /note= "Encoded by ACH"
FT   Misc-difference 36
FT                     /label= Arg, Ser
FT                     /note= "Encoded by MGM"
FT   Misc-difference 39
FT                     /label= Ala, Pro, Ser
FT                     /note= "Encoded by BCC"
FT   Misc-difference 40
FT                     /label= Arg, Gly, Trp
FT                     /note= "Encoded by BGG"
FT   Misc-difference 48
FT                     /label= Lys, Thr, Met, Glu, Ala, Val, Gln, Pro, Leu
FT                     /note= "Encoded by VHG"
FT   Misc-difference 54
FT                     /label= Val
FT                     /note= "Encoded by GTB"
FT   Misc-difference 56
FT                     /label= Lys, Glu, Gln
FT                     /note= "Encoded by VAG"
FT   Misc-difference 65
FT                     /label= Glu, Asp, Val
FT                     /note= "Encoded by GWS"
FT   Misc-difference 66
FT                     /label= Ile, Val, Leu
FT                     /note= "Encoded by VTC"
FT   Misc-difference 76
FT                     /label= Gly
FT                     /note= "Encoded by GGH"
FT   Misc-difference 87
FT                     /label= Thr, Pro, Ser
FT                     /note= "Encoded by HCB"
FT   Misc-difference 93
FT                     /label= Asp, Ala
FT                     /note= "Encoded by GMC"
FT   Misc-difference 102
FT                     /label= Lys, Asn
FT                     /note= "Encoded by AAV"
FT   Misc-difference 103
FT                     /label= Arg, Ser
FT                     /note= "Encoded by AGB"
FT   Misc-difference 109
FT                     /label= Thr, Ala, Ser
FT                     /note= "Encoded by DCB"
FT   Misc-difference 112
FT                     /label= Asp, His, Tyr
FT                     /note= "Encoded by BAC"
FT   Misc-difference 114
FT                     /label= Arg, Ser
FT                     /note= "Encoded by AGB"
FT   Misc-difference 117
FT                     /label= Gly
FT                     /note= "Encoded by GGH"
FT   Misc-difference 148
FT                     /label= Val
FT                     /note= "Encoded by GTB"
FT   Misc-difference 150
FT                     /label= Pro
FT                     /note= "Encoded by CCB"
FT   Misc-difference 153
FT                     /label= Ile, Val, Phe
FT                     /note= "Encoded by DTC"
FT   Misc-difference 154
FT                     /label= Glu, Asp
FT                     /note= "Encoded by GAB"
FT   Misc-difference 157
FT                     /label= Ser, Gly, Arg
FT                     /note= "Encoded by VGC"
FT   Misc-difference 159
FT                     /label= Gly
FT                     /note= "Encoded by GGH"
FT   Misc-difference 183
FT                     /label= Thr
FT                     /note= "Encoded by ACB"
FT   Misc-difference 223
FT                     /label= Pro
FT                     /note= "Encoded by CCB"
FT   Misc-difference 225
FT                     /label= Ser
FT                     /note= "Encoded by TCB"
FT   Misc-difference 237
FT                     /label= Ser, Thr, Ile, Gly, Ala, Val, Cys, Phe
FT                     /note= "Encoded by DVC"
FT   Misc-difference 239
FT                     /label= Val
FT                     /note= "Encoded by GTB"
FT   Misc-difference 241
FT                     /label= Val
FT                     /note= "Encoded by GTY"
FT   Misc-difference 242
FT                     /label= Val, Leu, Phe
FT                     /note= "Encoded by KTS"
FT   Misc-difference 246
FT                     /label= Lys, Asn, Thr
FT                     /note= "Encoded by AMS"
FT   Misc-difference 251
FT                     /label= Tyr, Ser, Phe
FT                     /note= "Encoded by THC"
FT   Misc-difference 259
FT                     /label= Leu, Phe
FT                     /note= "Encoded by TTD"
FT   Misc-difference 305
FT                     /label= Leu
FT                     /note= "Encoded by CTB"
FT   Misc-difference 319
FT                     /label= Pro
FT                     /note= "Encoded by CCB"
FT   Misc-difference 346
FT                     /label= Lys, Arg, Thr, Gln, Pro, Trp, Ser
FT                     /note= "Encoded by HVG or an in-frame stop codon"
FT   Misc-difference 411
FT                     /label= Lys, Thr, Met
FT                     /note= "Encoded by AHG"
FT   Misc-difference 421
FT                     /label= Thr, Ala, Ser
FT                     /note= "Encoded by DCC"
FT   Misc-difference 457
FT                     /label= Leu
FT                     /note= "Encoded by CTB"
FT   Misc-difference 458
FT                     /label= Pro
FT                     /note= "Encoded by CCB"
FT   Misc-difference 466
FT                     /label= Pro
FT                     /note= "Encoded by CCB"
FT   Misc-difference 487
FT                     /label= Gly
FT                     /note= "Encoded by GGH"
FT   Misc-difference 493
FT                     /label= Val
FT                     /note= "Encoded by GTB"
FT   Misc-difference 500
FT                     /note= "Encoded by GCH"
FT   Misc-difference 518
FT                     /label= Thr
```

```
FT                    /note= "Encoded by ACB"
FT   Misc-difference 536
FT                    /label= Val
FT                    /note= "Encoded by GTB"
FT   Misc-difference 556
FT                    /label= Met, Ile, Val, Leu
FT                    /note= "Encoded by VTB"
FT   Misc-difference 557
FT                    /label= Ile, Val, Leu
FT                    /note= "Encoded by VTC"
FT   Misc-difference 558
FT                    /label= Thr, Ala, Pro
FT                    /note= "Encoded by VCB"
FT   Misc-difference 559
FT                    /label= Thr, Ala, Ser
FT                    /note= "Encoded by DCG"
FT   Misc-difference 561
FT                    /label= Ile, Leu, Phe
FT                    /note= "Encoded by HTC"
FT   Misc-difference 562
FT                    /label= Lys, Glu, Gln
FT                    /note= "Encoded by VAG"

  Query Match             75.6%;  Score 2730;  DB 5;  Length 720;
  Best Local Similarity   76.5%;  Pred. No. 1.2e-229;
  Matches  551;  Conservative   45;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSXDSIATSXANGXXNXXHAAPXKXXTGATSXLXRPXXLAPPATQXTQLDIXEXILAD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         61 PTDDXXELDGYTLTLXDVVGAARKGRXVRVQTXDEIRAKIDXXVEFLRXQLXNXVYXVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        121 GFGGSADTRTEDAISLQKALLEHQLCGXLXTSXXSFXLXRGLENSLPLEVVRGAMTIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        181 SLXRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSXLXYIAAAITGHPDXKXH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        241 XXHEGXEKIMXAREAIALXGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        301 ALTAXTVEAMVGHAGSFHXFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
Db        361 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAESTTDNPLIDVENKXTHHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
Db        421 XAVANTMEKTRLALALIGKLNFTQLTEMLNAGMNRGXXSCLAAEDXSLSYHCKGLDIAAA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
Db        481 AYTSELXHLANPXTTHVQPREMGNQAVNSLALISARRXAEANDVLSLLLATHLYCXLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
Db        541 DLRAMEFEFKKQFDPXXXXLXXQHFGXALDGXELXDKVNKXLXKRLEQTNSYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
Db        601 AFSFATGTVVELLSSSXXAKVSLAAVNAWKVASAEKAISLTRXVRDXFWXAPSSSSPALX 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
Db        661 YLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGTNVSRIYEAIKXGXINHVLVKMLA 720

RESULT 6
AAP83141
ID   AAP83141 standard; protein; 716 AA.
XX
AC   AAP83141;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   19-NOV-1990  (first entry)
XX
DE   Sequence of L-phenylalanine ammonia lyase (PAL) from Rhodosporidium
DE   turoloides.
XX
KW   Cinnamic acid; expression vector.
XX
OS   Rhodosporidium toruloides.
XX
PN   EP278706-A.
XX
PD   17-AUG-1988.
XX
PF   08-FEB-1988;   88EP-00301011.
XX
PR   06-FEB-1987;   87JP-00024705.
PR   18-JUN-1987;   87JP-00152357.
XX
PA   (MITK ) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N,  Yoshino S,  Sone S,  Nakajima Y,  Makiguchi N;
XX
DR   WPI; 1988-229543/33.
XX
PT   Recombinant plasmid for expression of L-phenylalanine ammonia-lyase -
PT   having combined promoter of tac promoter and Pl promoter for enhanced
PT   expression.
XX
PS   Example; Page 16-21; 37pp; English.
XX
CC   Inventors claim a recombinant plasmid contg. a DNA sequence coding for
CC   PAL with the SQ in AAP80501. Procedure for cloning the structural gene
CC   for PAL (AAN81101) is described as a reference example. The hybrid
CC   plasmid permits more efficient expression of PAL in E.coli. PAL is used
CC   for the prodn. of L-phenylalanine from cinnamic acid and ammonia (Updated
CC   on 25-MAR-2003 to correct PR field.) (Updated on 27-AUG-2003 to correct
CC   OS field.)
XX
SQ   Sequence 716 AA;

  Query Match             72.5%;  Score 2619;  DB 1;  Length 716;
  Best Local Similarity   72.1%;  Pred. No. 6.2e-220;
  Matches  523;  Conservative   86;  Mismatches  102;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
```

```
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :||:|||||||   ||| ||:|||  ||||  |:|||  |||  ||  ||   |||   |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:|||||  |::||::  |::|     :|||||:|||||  :  |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||  ||  |||||||  ||:|||  ||||||||||||||||:||  |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              |||  ||||||||||||  ||||||  ||||||||||||||||  |:|||||||||:||||  |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:|  :|||||  |  :  :|:  ||  |     ||  :: :  |    :|  |  |  ||||||:|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTTSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:|  ||||||:|||:      :::|  |||||||||:||  ||||||:||   ||     |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :||  :|||||||::||:||||||||||:|||||||:|  |:  |||||||:||||||  |||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 7
AAP80513
ID AAP80513 standard; protein; 716 AA.
XX
AC AAP80513;
XX
DT 25-MAR-2003 (revised)
DT 12-NOV-1990 (first entry)
XX
DE L-phenylalanine ammonialyase.
XX
KW L-phenylalanine ammonialyase; PAL; foreign gene expression;
KW culture temperature; expression regulation.
XX
OS Rhodosporidium toruloides.
XX
PN EP279665-A.
XX
PD 24-AUG-1988.
XX
PF 18-FEB-1988; 88EP-00301356.
XX
PR 19-FEB-1987; 87JP-00034397.
PR 18-JUN-1987; 87JP-00152359.
XX
PA (MITK) MITSUI TOATSU CHEM INC.
XX
PI Fukuhara N, Yoshino S, Yamamoto K, Sone S, Suzuki M, Nakajima Y;
XX
DR WPI; 1988-236895/34.
DR N-PSDB; AAN81116.
XX
PT Regulation of expression of foreign gene in Escherichia coli - by
PT maintaining temp. of culture at 40 deg. C or more to suppress expression.
XX
PS Claim 3; Page 15-19; 36pp; English.
XX
CC The PAL gene is expressed in E.coli carrying a recombinant plasmid.
CC Expression is regulated by maintaining the temperature at at least 40
CC deg. C to suppress expression. Pal is produced in high concentrations.
CC See also EP-279664: regualtion by sugar component as C source at 0.3% or
CC more. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
CC 2003 to correct PI field.)
XX
SQ Sequence 716 AA;

```
  Query Match             72.4%;  Score 2614;  DB 1;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.7e-219;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: :  |||            |   |   |       |          |          :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :||  ||||  :  |||  ||||||  |||::  ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:|||||||:||||||  |:  ||  :|||||:|||||||||||  ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||  ||||||||:||||||||||||||  ||:||||  |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:|||  |||::||||::||  ||  ||||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||   |||  ||:|||  ||||  |:|||  |||  ||  ||   |||  |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:|||||  |::||::  |::|     :|||||:|||||  :  |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||  ||  |||||||  ||:|||  ||||||||||||||||:||  |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              |||  ||||||||||||  ||||||  ||||||||||||||||  |:|||||||||:||||  |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:|  :|||||  |  :  :|:  ||  |     ||  :: :  |    :|  |  |  ||||||  :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:|  ||||||:|||:      :::|  |||||||||:||  ||||||:||   ||     |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :||  :|||||||::||:||||||||||:|||||||:|  |:  |||||||:||||||  |||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 8
ABB07693
ID ABB07693 standard; protein; 716 AA.
XX
AC ABB07693;
XX
DT 07-AUG-2003 (revised)
DT 10-JUN-2002 (first entry)

```
XX
DE   R. toruloides PAL polypeptide.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; enzyme;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
OS   Rhodosporidium toruloides.
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   N-PSDB; ABA95243.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Example 3; Fig 1; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a R. toruloides PAL polypeptide. (Updated on 07-AUG-2003 to
CC   correct OS field.)
XX
SQ   Sequence 716 AA;

  Query Match             72.4%;  Score 2614;  DB 5;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.7e-219;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||         |  |     |         |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| ||||  : ||| |||||| |||:: ||||:::||||:||::|| |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| |||||||||:|||||||||||||||  ||:||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:|||||  |::||::  |::|      :|||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || ||||||| ||:||| ||||||||||||||||:|| ||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| |||||| ||||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :|||||  | : :|: ||  |     || :: : |  :|  |  |  ||||||  :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:   :::| ||||||||:|| ||||||:||  ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716

RESULT 9
AAE27941
ID   AAE27941 standard; protein; 716 AA.
XX
AC   AAE27941;
XX
DT   06-AUG-2003  (revised)
DT   27-DEC-2002  (first entry)
XX
DE   Rhodotorula toruloides PAL protein.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide.
XX
OS   Rhodosporidium toruloides.
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   N-PSDB; AAD45815.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Example 3; Page 43-45; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
```

```
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is R. toruloides PAL
CC   protein. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 716 AA;

  Query Match             72.4%;  Score 2614;  DB 5;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.7e-219;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
                |||||||:: : |||          |  | |   |      |          :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||  ||| ||:||| |||| |:||| |||  ||| || || ||| ||||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||||:||||| |::||:: |::| |||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || ||||||||| ||:||| |||||||||||||||:|| ||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||||| |||||| |||||||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :|||||| | : :|: || |   || :: : :| | | ||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:   :::| ||||||||:|| ||||||:|| ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:|||||||||:|||||||:| |: |||||||:||||||  |||:|:|
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716

RESULT 10
AAE20653
ID   AAE20653 standard; protein; 716 AA.
XX
AC   AAE20653;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Rhodotorula glutinis phenylalanine ammonia lyase (PAL).
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst.
XX
OS   Rhodotorula glutinis.
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   N-PSDB; AAD33075.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Example 6; Page 92-95; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL protein
XX
SQ   Sequence 716 AA;

  Query Match             72.4%;  Score 2614;  DB 5;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.7e-219;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
                |||||||:: : |||          |  | |   |      |          :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
```

```
               |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db         235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
               :|||:||||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db         295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
               ||||||||||||||||:||||| |::||:: |::|  :||||||:||||| : |||||
Db         355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy         420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
               |:||: :|||||| || ||||||||| ||:||| ||||||||||||||||:|| |||||
Db         415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy         480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
               ||| |||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db         475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy         540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
               :||||:| :|||||  | : :|: || |     || :: : |  : |  | ||||| :|||
Db         535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy         596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
               ||||||||:| |||||:|||:     :::| ||||||||:|| ||||||:|| ||    |:
Db         595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy         656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
               :|| :||||||::||:|||||||||:||||||||:| |: ||||||||:|||||| ||||:||
Db         652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy         716 VKMLA 720
               :||||
Db         712 LKMLA 716
```

RESULT 11
AAE16389
ID   AAE16389 standard; protein; 716 AA.
XX
AC   AAE16389;
XX
DT   09-APR-2002 (first entry)
XX
DE   Rhodotorula glutinis wildtype phenylalanine ammonia-lyase (PAL).
XX
KW   Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
KW   para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW   Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW   aerospace application; chemical; red yeast.
XX
OS   Rhodotorula glutinis.
XX
PN   WO200111071-A2.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US021156.
XX
PR   06-AUG-1999;   99US-0147719P.
XX
PA   (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI   Tang X, Vannelli TM, Qi WW, Sariaslani S, Gatenby AA;
XX
DR   WPI; 2002-121549/16.
DR   N-PSDB; AAD26923.
XX
PT   Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT   comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia
PT   -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT   activity.
XX
PS   Claim 8; Page 65-67; 75pp; English.
XX
CC   The invention relates to a method for producing para-hydroxycinnamic acid
CC   (PHCA). The method comprises converting cinnamate to PHCA by converting
CC   glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC   route or by generating a new biocatalyst possessing enhanced tyrosine
CC   ammonia-lyase (TAL) activity. The method is useful for the biological
CC   production of PHCA which is useful as a monomer for the production of
CC   Liquid Crystal Polymers (LCP), where the LCP is used in electronic
CC   connectors, telecommunications and aerospace applications. LCP is also
CC   useful in medical devices, as well as chemical, and food packing
CC   applications due to its resistance to sterilising radiation. The present
CC   sequence is red yeast, Rhodotorula glutinis (Rhodosporidium toruloides)
CC   wildtype PAL
XX
SQ   Sequence 716 AA;

```
  Query Match             72.4%;  Score 2614;  DB 5;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.7e-219;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy           1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
               |||||||:: : |||          |  |  |      |        |             :|::||:::|:
Db           1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy          61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
               |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::|| ||||||
Db          55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy         121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
               |||||||||||:||||||:||||||| |:|  || :||||||:|||||||||| ||||
Db         115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy         181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
               |||||||||||||||||||||||||:||||||||||||||||||||| ||:|||| |||
Db         175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy         241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
               |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db         235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
               :|||:||||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db         295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
               ||||||||||||||||:||||| |::||:: |::|  :||||||:||||| : |||||
Db         355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy         420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
               |:||: :|||||| || ||||||||| ||:||| ||||||||||||||||:|| |||||
Db         415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy         480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
               ||| |||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db         475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy         540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
               :||||:| :|||||  | : :|: || |     || :: : |  : |  | ||||| :|||
Db         535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy         596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
               ||||||||:| |||||:|||:     :::| ||||||||:|| ||||||:|| ||    |:
Db         595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy         656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
               :|| :||||||::||:|||||||||:||||||||:| |: ||||||||:|||||| ||||:||
Db         652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy         716 VKMLA 720
               :||||
Db         712 LKMLA 716
```

```
RESULT 12
ABG71948
ID   ABG71948 standard; protein; 716 AA.
XX
AC   ABG71948;
XX
DT   27-JAN-2003  (first entry)
XX
DE   R. glutinis wild-type Phenylalanine ammonia-lyase, PAL.
XX
KW   Enzyme; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase; PAH;
KW   para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; tyrosine;
KW   cinnamic acid; PHCA; liquid crystal polymer.
XX
OS   Rhodtorula glutinis.
XX
PN   WO200290523-A2.
XX
PD   14-NOV-2002.
XX
PF   03-MAY-2002; 2002WO-US018551.
XX
PR   04-MAY-2001; 2001US-0288701P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Qi WW,  Sariaslani FS,  Tang X;
XX
DR   WPI; 2003-058928/05.
DR   N-PSDB; ABS56530.
XX
PT   Novel recombinant host for producing para-hydroxycinnamic acid and
PT   tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT   gene encoding phenylalanine hydroxylase activity.
XX
PS   Claim 5; Page 48-50; 69pp; English.
XX
CC   The invention relates to a recombinant host comprising at least one gene
CC   encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
CC   encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
CC   (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC   providing the recombinant host; and (2) growing the recombinant organism
CC   in the presence of a fermentable carbon substrate and thus producing para
CC   -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC   providing a recombinant organism comprising at least one gene encoding a
CC   phenylalanine hydroxylase activity; and (2) growing the recombinant
CC   organism in the presence of a fermentable carbon substrate to produce
CC   tyrosine. The recombinant host cell is used for producing para-
CC   hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC   production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC   monomer for the production of liquid crystal polymer. The present
CC   sequence is R. glutinis wild-type Phenylalanine ammonia-lyase, PAL, an
CC   enzyme which catalyses the conversion of phenylalanine into cinnamic acid
XX
SQ   Sequence 716 AA;

  Query Match             72.4%;  Score 2614;  DB 6;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.7e-219;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||       |  |  |     |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| ||||  : ||| |||||| |||:: ||||:::||||:||::||  |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||| |||||||||:||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||   ||| ||:||| |||| |:||| ||| || |||  |||||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:|||| |::||:: |::|   :||||||:|||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || ||||||||| ||:||| |||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :|||||  | : :|: || |    || :: :  :| |  | ||||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              |||||||:| |||||:|||:    :::| ||||||||:|| |||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:||||||||:| |: ||||||:||||||  ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716


RESULT 13
ADE36054
ID   ADE36054 standard; protein; 716 AA.
XX
AC   ADE36054;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rhodosporidium glutinis PAL protein SEQ ID NO:2.
XX
KW   recombination; nucleic acid template; recombined polypeptide;
KW   altered property; combined gene.
XX
OS   Rhodosporidium.
XX
PN   WO2003072743-A2.
XX
PD   04-SEP-2003.
XX
PF   26-FEB-2003; 2003WO-US005708.
XX
PR   26-FEB-2002; 2002US-0360279P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Milano J,  Tang X;
XX
DR   WPI; 2003-721766/68.
DR   N-PSDB; ADE36053.
XX
PT   Recombination of nucleic acid templates, useful e.g. for producing
PT   enzymes with altered properties, is based on cyclic extension of unpaired
```

```
PT   primers.
XX
PS   Disclosure; SEQ ID NO 2; 481pp; English.
XX
CC   The present invention describes a method for the recombination of nucleic
CC   acid (NA) templates. The method comprises: (a) providing at least two
CC   double-stranded templates (T1, T2) having different 5' and 3' regions in
CC   their sense strands; (b) contacting the templates with at least one each
CC   of forward and reverse primers that anneal, respectively, only to the 3'-
CC   region of the antisense strand of T1 and to the 3'-region of the sense
CC   strand of T2; (c) extending the primers by no more than 1000 nucleotides
CC   (nt); (d) separating extended primers from their templates; (e)
CC   reannealing the extended primers, to either template; and (f) repeating
CC   cycles of steps (b)-(e) until at least one full-length extension product,
CC   i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
CC   method starting from at least one each of antisense and sense single-
CC   stranded templates; (2) generating a recombined polypeptide (I) with
CC   altered properties by expressing recombination products of the new
CC   processes and screening the polypeptides formed for properties different
CC   from those of polypeptides from either of the templates; and (3) (I)
CC   produced by method (2). The methods are useful for the recombination of
CC   nucleic acid templates of interest, and generation of a recombined
CC   polypeptide having altered properties. The method can be used to make
CC   combined genes that express proteins e.g. enzymes, cytokines, growth
CC   factors, viral proteins or microbial antigens with altered properties,
CC   e.g. stability, activity or specificity. The present sequence is used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 716 AA;

  Query Match             72.4%;  Score 2614;  DB 7;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.7e-219;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||         |  |  |     |       |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| |||| : ||| |||||| |||:: ||||:::|||:||::|| |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||||:||||||:|||||||| |:|  || :||||||:||||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||| ||| ||:|||  ||||  |:||| |||  ||  |||  |||||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:|||| |::||:: |::|   :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || ||||||| ||:||| |||||||||||||||||:|| ||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| |||||||||||||| :||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :||||| | : :|: || |   || :: : |  :| | | |||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:   :::| ||||||||:|| ||||||:|| ||  |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||||::||:|||||||:|||||||:| |: ||||||:|||||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

```
RESULT 14
AAE20660
ID   AAE20660 standard; protein; 716 AA.
XX
AC   AAE20660;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Rhodotorula glutinis mutant tyrosine ammonia lyase (TAL), RM120-2 #1.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; mutant; variant; mutein.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 198
FT                   /label= Asp, Asn, Glu, Gln
FT   Misc-difference 540
FT                   /label= Thr, Ala, Ser, Pro, Gly
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Claim 2; Page 126-128; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
```

```
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   mutant TAL
XX
SQ   Sequence 716 AA;

  Query Match             72.4%;  Score 2613;  DB 5;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 2.1e-219;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |     |       |      :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||||  |:|   ||  :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||| ||| ||:|| |||  |:||| ||| || ||| ||  |||  |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||:||||| |::|:: |::|   :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| ||||| |||||||||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :||||| | : :|: || |   || :: : |  :| | | ||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:   :::| ||||||||:|| ||||||:|| ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||::||:||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

```
RESULT 15
ABG71951
ID   ABG71951 standard; protein; 716 AA.
XX
AC   ABG71951;
XX
DT   27-JAN-2003  (first entry)
XX
DE   R. glutinis PAL/TAL mutant #3.
XX
KW   Enzyme; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase; PAH;
KW   para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
KW   tyrosine; cinnamic acid; PHCA; liquid crystal polymer; mutant; mutein.
XX
OS   Rhodtorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 198
FT                   /label= Asp, Asn, Glu, Gln
FT   Misc-difference 540
FT                   /label= Thr, Ala, Ser, Pro, Gly
XX
PN   WO200290523-A2.
XX
PD   14-NOV-2002.
XX
PF   03-MAY-2002; 2002WO-US018551.
XX
PR   04-MAY-2001; 2001US-0288701P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Qi WW,  Sariaslani FS,  Tang X;
XX
DR   WPI; 2003-058928/05.
XX
PT   Novel recombinant host for producing para-hydroxycinnamic acid and
PT   tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT   gene encoding phenylalanine hydroxylase activity.
XX
PS   Claim 5; Page 57-59; 69pp; English.
XX
CC   The invention relates to a recombinant host comprising at least one gene
CC   encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
CC   encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
CC   (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC   providing the recombinant host; and (2) growing the recombinant organism
CC   in the presence of a fermentable carbon substrate and thus producing para
CC   -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC   providing a recombinant organism comprising at least one gene encoding a
CC   phenylalanine hydroxylase activity; and (2) growing the recombinant
CC   organism in the presence of a fermentable carbon substrate to produce
CC   tyrosine. The recombinant host cell is used for producing para-
CC   hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC   production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC   monomer for the production of liquid crystal polymer. The present
CC   sequence is a mutant of the R. glutinis Phenylalanine ammonia-lyase, PAL,
CC   which has TAL activity (PAL/TAL)
XX
SQ   Sequence 716 AA;

  Query Match             72.4%;  Score 2613;  DB 6;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 2.1e-219;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |     |       |      :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||||  |:|   ||  :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234
```

```
Qy     241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
           |:||| |||::||||::|| || |||||||||||||||||||||||||:|||:|||||||
Db     235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy     301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
           :|||:|||||||   ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db     295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy     361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
           |||||||||||||||:||||| |::||:: |::|    :||||||:||||| : ||||||
Db     355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy     420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
           |:||: :||||| || ||||||| ||:||| |||||||||||||||||:|| |||||
Db     415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy     480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
           ||| ||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db     475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy     540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
           :|||| | :||||| | : :|: || |     || :: : |  :| | | |||||| :|||
Db     535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy     596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
           ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||   |:
Db     595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy     656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
           :|| :||||||::||:||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db     652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy     716 VKMLA 720
           :||||
Db     712 LKMLA 716
```

Search completed: September 9, 2004, 10:04:00
Job time : 92.6307 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:          September  9, 2004, 10:01:03 ; Search time 23.4025 Seconds
                                              (without alignments)
                                              1588.322 Million cell updates/sec

Title:           US-09-939-408A-13
Perfect score:   3610
Sequence:        1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3608	99.9	720	4	US-09-624-693A-13	Sequence 13, Appl
2	2770	76.7	726	4	US-09-624-693A-21	Sequence 21, Appl
3	2614	72.4	716	4	US-09-624-693A-19	Sequence 19, Appl
4	2614	72.4	716	4	US-09-627-216A-8	Sequence 8, Appli
5	2614	72.4	716	4	US-09-765-873A-8	Sequence 8, Appli
6	2613	72.4	716	4	US-09-765-873A-35	Sequence 35, Appl
7	2612	72.4	716	4	US-09-627-216A-10	Sequence 10, Appl
8	2612	72.4	716	4	US-09-765-873A-10	Sequence 10, Appl
9	2607	72.2	716	4	US-09-765-873A-38	Sequence 38, Appl
10	2603	72.1	716	4	US-09-765-873A-36	Sequence 36, Appl
11	2603	72.1	716	4	US-09-765-873A-37	Sequence 37, Appl
12	2599	72.0	716	4	US-09-765-873A-34	Sequence 34, Appl
13	2579	71.4	716	4	US-09-765-873A-33	Sequence 33, Appl
14	2573	71.3	686	4	US-09-765-873A-32	Sequence 32, Appl
15	2499.5	69.2	713	4	US-09-624-693A-17	Sequence 17, Appl
16	1346.5	37.3	740	4	US-09-624-693A-15	Sequence 15, Appl
17	716	19.8	424	4	US-09-615-192A-328	Sequence 328, App
18	524	14.5	525	4	US-09-540-236-2855	Sequence 2855, Ap
19	508.5	14.1	515	4	US-09-328-352-5090	Sequence 5090, Ap
20	503.5	13.9	508	4	US-09-489-039A-13575	Sequence 13575, A
21	500.5	13.9	515	4	US-09-252-991A-33046	Sequence 33046, A
22	486	13.5	412	4	US-09-615-192A-330	Sequence 330, App
23	397	11.0	520	4	US-09-252-991A-33049	Sequence 33049, A
24	241	6.7	226	4	US-09-615-192A-327	Sequence 327, App
25	211.5	5.9	97	4	US-09-615-192A-329	Sequence 329, App
26	129	3.6	2123	3	US-08-968-685A-10	Sequence 10, Appl
27	127	3.5	2588	3	US-08-936-135-2	Sequence 2, Appli
28	124	3.4	857	4	US-09-252-991A-23956	Sequence 23956, A
29	118.5	3.3	1832	3	US-09-335-409-4	Sequence 4, Appli
30	118.5	3.3	1832	4	US-09-568-102-4	Sequence 4, Appli
31	118.5	3.3	1832	4	US-09-567-969-4	Sequence 4, Appli
32	118.5	3.3	1832	4	US-09-568-480-4	Sequence 4, Appli
33	118.5	3.3	1832	4	US-09-568-486-4	Sequence 4, Appli
34	118.5	3.3	1832	4	US-09-568-472-4	Sequence 4, Appli
35	118.5	3.3	1832	4	US-09-567-899-4	Sequence 4, Appli
36	118	3.3	647	4	US-09-613-303-53	Sequence 53, Appl
37	118	3.3	647	4	US-10-267-311-53	Sequence 53, Appl
38	117	3.2	132	4	US-09-615-192A-331	Sequence 331, App
39	117	3.2	1839	2	US-09-172-977-4	Sequence 4, Appli
40	117	3.2	1839	4	US-09-404-108-4	Sequence 4, Appli
41	116.5	3.2	1239	1	US-08-026-138E-3	Sequence 3, Appli
42	116.5	3.2	1589	4	US-09-543-681A-4998	Sequence 4998, Ap
43	115.5	3.2	366	4	US-09-252-991A-30639	Sequence 30639, A
44	115.5	3.2	1214	2	US-08-231-193A-54	Sequence 54, Appl
45	115.5	3.2	1214	2	US-08-486-273A-54	Sequence 54, Appl

ALIGNMENTS

```
RESULT 1
US-09-624-693A-13
; Sequence 13, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSC
;  CURRENT APPLICATION NUMBER: US/09/624,693A
;  CURRENT FILING DATE:  2000-07-24
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Rhodotorula graminis
US-09-624-693A-13

  Query Match             99.9%;  Score 3608;  DB 4;  Length 720;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  719;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSVSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
```

```
Db      361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420

Qy      421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
Db      421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480

Qy      481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
Db      481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540

Qy      541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
Db      541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600

Qy      601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
Db      601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660

Qy      661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
Db      661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
```

RESULT 2
US-09-624-693A-21
; Sequence 21, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12, 16, 17, 20, 24, 25, 28 - 36, 38, 42, 47, 48, 56, 57, 62, 66,
; OTHER INFORMATION: "Xaa" means any amino acid; "Xaa" means no consensus at that
; OTHER INFORMATION: position
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOs: 13, 17, and 19
US-09-624-693A-21

```
  Query Match             76.7%;  Score 2770;  DB 4;  Length 726;
  Best Local Similarity   77.7%;  Pred. No. 1.3e-262;
  Matches  564;  Conservative   40;  Mismatches  116;  Indels    6;  Gaps    3;

Qy        1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db        1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy       61 P-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVT 119
Db       61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy      120 TGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRV 179
Db      121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy      180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239
Db      181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy      240 HVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
Db      241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy      300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
Db      301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy      360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNF 418
Db      361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy      419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIH 478
Db      421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy      479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
Db      481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy      539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA----LALEVKKALNKRLEQTTTYDL 594
Db      541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy      595 EPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSS 654
Db      601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy      655 QAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHV 714
Db      661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy      715 LVKMLA 720
Db      721 LVKMLA 726
```

RESULT 3
US-09-624-693A-19
; Sequence 19, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula toruloides
US-09-624-693A-19

```
  Query Match             72.4%;  Score 2614;  DB 4;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 2.6e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy        1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db        1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy       61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db       55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy      121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db      115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174
```

```
Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 4
US-09-627-216A-8
; Sequence 8, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-09-627-216A-8

Query Match 72.4%; Score 2614; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 2.6e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```
Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 5
US-09-765-873A-8
; Sequence 8, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-09-765-873A-8

Query Match 72.4%; Score 2614; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 2.6e-247;

```
  Matches  522;  Conservative  86;  Mismatches  103;  Indels  14;  Gaps  5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |    |       |       :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| |||| : ||| |||||| |||::  ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||:||||||:|||||| |:|   || :|||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||| |||||||||:||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || |||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||  |||  ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:||||| |::||:: |::|    :|||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| ||||| |||||||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :||||| |  : :|: || |    || :: : |    :| | | ||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:|||||||||:||||||||:| |: ||||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 6
US-09-765-873A-35
; Sequence 35, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: X= Asp, Asn, Glu, Gln
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-35

```
  Query Match             72.4%;  Score 2613;  DB 4;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 3.2e-247;
  Matches  522;  Conservative  85;  Mismatches  104;  Indels  14;  Gaps  5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |    |       |       :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| |||| : ||| |||||| |||::  ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||:||||||:|||||| |:|   || :|||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||| |||||||||:||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || |||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||  |||  ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:||||| |::||:: |::|    :|||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| ||||| |||||||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :||||| |  : :|: || |    || :: : |    :| | | ||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:|||||||||:||||||||:| |: ||||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 7
US-09-627-216A-10
; Sequence 10, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:

```
;  APPLICANT: Sariaslani, Sima F
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  APPLICANT:  Vannelli, Todd
;  APPLICANT:  Gatenby, Anthony
;  TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
;  FILE REFERENCE: BC1009 US NA
;  CURRENT APPLICATION NUMBER: US/09/627,216A
;  CURRENT FILING DATE:  2000-07-27
;  PRIOR APPLICATION NUMBER: 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: mutant from
;   OTHER INFORMATION: Rhodtorula glutinis
US-09-627-216A-10

  Query Match             72.4%;  Score 2612;  DB 4;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 4e-247;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy       1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
           |||||||:: : |||       |  | |    |      |        :|::||:::|:
Db       1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy      61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
           |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  |||||||
Db      55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy     121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
           |||||||||||||:||||||:||||||| |:|  || :||||||:|||||||||| ||||
Db     115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy     181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
           ||||||||||||||||||||||| |||||||||:|||||||||||||| ||:|||| |||
Db     175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy     241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
           |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db     235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy     301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
           :|||:|||||||  ||| ||:||| |||| |:||| ||| || || ||| ||||||||||
Db     295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy     361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
           ||||||||||||||||:||||| |::||:: |::|   :||||||:||||| : ||||||
Db     355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy     420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
           |:||: :|||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db     415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy     480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
           ||| |||||||||||| |||||| ||||||||||||||| |:|||||||||:|||| |||
Db     475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy     540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
           :|||| | :||||| | : :|: || |   || :: : |  :| | | |||||| :|||
Db     535 IDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy     596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
           ||||||||:| |||||:|||:   :::| ||||||||:|| ||||||:||  ||   |:
Db     595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy     656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
           :|| :||||||::||:||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db     652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy     716 VKMLA 720
           :||||
Db     712 LKMLA 716


RESULT 8
US-09-765-873A-10
; Sequence 10, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-10

  Query Match             72.4%;  Score 2612;  DB 4;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 4e-247;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy       1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
           |||||||:: : |||       |  | |    |      |        :|::||:::|:
Db       1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy      61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
           |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  |||||||
Db      55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy     121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
           |||||||||||||:||||||:||||||| |:|  || :||||||:|||||||||| ||||
Db     115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy     181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
           ||||||||||||||||||||||| |||||||||:|||||||||||||| ||:|||| |||
Db     175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy     241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
           |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db     235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy     301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
           :|||:|||||||  ||| ||:||| |||| |:||| ||| || || ||| ||||||||||
Db     295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy     361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
           ||||||||||||||||:||||| |::||:: |::|   :||||||:||||| : ||||||
Db     355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy     420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
           |:||: :|||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db     415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy     480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
           ||| |||||||||||| |||||| ||||||||||||||| |:|||||||||:|||| |||
Db     475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy     540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
           :|||| | :||||| | : :|: || |   || :: : |  :| | | |||||| :|||
Db     535 IDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy     596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
```

```
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:||||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716


RESULT 9
US-09-765-873A-38
; Sequence 38, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 38
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

  Query Match              72.2%;  Score 2607;  DB 4;  Length 716;
  Best Local Similarity    71.9%;  Pred. No. 1.3e-246;
  Matches  521;  Conservative   85;  Mismatches  105;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||          |  |  |     |     |         :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::|| ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||| |||||||||:||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::|||::|| || ||||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:||||| |::||:: |::|    :|||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || |||||||| ||:||| |||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||||| |||||| ||| |||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :|||||| | : :|: || |     || :: : | :|  | | |||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:||||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716


RESULT 10
US-09-765-873A-36
; Sequence 36, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (181)..(181)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (235)..(235)
;   OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-36

  Query Match              72.1%;  Score 2603;  DB 4;  Length 716;
  Best Local Similarity    71.7%;  Pred. No. 3.1e-246;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||          |  |  |     |     |         :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::|| ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174
```

```
Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHXAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 XVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 11
US-09-765-873A-37
; Sequence 37, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (149)..(149)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (202)..(202)
;   OTHER INFORMATION: X= Val, Met, Leu, Cys
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-37

```
  Query Match             72.1%;  Score 2603;  DB 4;  Length 716;
  Best Local Similarity   71.7%;  Pred. No. 3.1e-246;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDXFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPXVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 12
US-09-765-873A-34
; Sequence 34, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97

```
; SEQ ID NO 34
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (126)..(126)
;   OTHER INFORMATION: X= Gly, Ala, Ser, Thr
;   NAME/KEY: UNSURE
;   LOCATION: (138)..(138)
;   OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-34

  Query Match             72.0%;  Score 2599;  DB 4;  Length 716;
  Best Local Similarity   71.7%;  Pred. No. 7.6e-246;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||          | | |    |       |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| ||||  : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||| |:||||||:|| |||| |:|   || :|||||:|||||||||||| ||||
Db        115 GFGGSADTRTEXAISLQKALLEHXLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| ||||||||:|||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| ||.||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||   ||| ||:|||  ||| |:|||  ||| |  |||  |||||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:|||||  |::||:: |::|     :||||||:|||||  : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||||  ||  |||||||| ||:|||  ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| ||||||  ||||||||||||||| |:||||||||||:|||: ||||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||  | :||||  |  : :|: ||  |   || :: :  |  :|  |  |||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              |||||||:| |||||:|||:     :::| ||||||||:|| ||||||:|| ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||||::||:|||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

```
RESULT 13
US-09-765-873A-33
; Sequence 33, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (126)..(126)
;   OTHER INFORMATION: X= Gly, Ser, Ala, Thr
;   NAME/KEY: UNSURE
;   LOCATION: (138)..(138)
;   OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
;   NAME/KEY: UNSURE
;   LOCATION: (149)..(149)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (181)..(181)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: X= Asp, Asn, Glu, Gln
;   NAME/KEY: UNSURE
;   LOCATION: (202)..(202)
;   OTHER INFORMATION: X= Val, Met, Leu, Cys
;   NAME/KEY: UNSURE
;   LOCATION: (235)..(235)
;   OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-09-765-873A-33

  Query Match             71.4%;  Score 2579;  DB 4;  Length 716;
  Best Local Similarity   71.0%;  Pred. No. 7e-244;
  Matches  515;  Conservative   86;  Mismatches  110;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||          | | |    |       |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| ||||  : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||| |:||||||:|| |||| |:|   | :||||||:|||||||||||| ||||
Db        115 GFGGSADTRTEXAISLQKALLEHXLCGVLPSSFDXFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||| ||||||||||||||| ||| |||||:|||||||||||||| ||:|||| |||
Db        175 SLTRGHXAVRLVVLEALTNFLNHXITPXVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              :||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db        235 XVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294
```

```
Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 14
US-09-765-873A-32
; Sequence 32, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 686
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-32

```
  Query Match             71.3%;  Score 2573;  DB 4;  Length 686;
  Best Local Similarity   74.9%;  Pred. No. 2.5e-243;
  Matches  507;  Conservative   83;  Mismatches   79;  Indels    8;  Gaps    4;

Qy         49 SQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLK 108
Db         13 TQVDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLR 72

Qy        109 AQLQNSVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPL 168
Db         73 SQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPL 132

Qy        169 EVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIA 228
Db        133 EVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIA 192

Qy        229 GAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLS 288
Db        193 AAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLA 252

Qy        289 LHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFA 348
Db        253 LHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFA 312

Qy        349 VEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDV 407
Db        313 VHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDV 372

Qy        408 ENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPS 467
Db        373 ENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPS 432

Qy        468 LNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSL 527
Db        433 LSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSL 492

Qy        528 LLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALN 583
Db        493 LLATHLYCVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLA 552

Qy        584 KRLEQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTRE 643
Db        553 KRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQ 609

Qy        644 VRNRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIY 703
Db        610 VRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIY 669

Qy        704 EAIKDGRINHVLVKMLA 720
Db        670 EAIKSGRINNVLLKMLA 686
```

RESULT 15
US-09-624-693A-17
; Sequence 17, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Rhodotorula mucilaginosa
US-09-624-693A-17

```
  Query Match             69.2%;  Score 2499.5;  DB 4;  Length 713;
  Best Local Similarity   69.6%;  Pred. No. 4.5e-236;
  Matches  511;  Conservative   83;  Mismatches  105;  Indels   35;  Gaps    9;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPT-------KSAAGPTSALRRTPGLDGHAAHQSQLEI 53
Db          1 MAPSVDSIATSVANSLSNGLHAAAAANGGDVHKKTAGAGSLLPTT--------ETTQLDI 52

Qy         54 VQELLSDP-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQ 112
Db         53 VERILADAGATDQIKLDGYTLTLGDVVGAARRGRSVKVADSPHIREKIDASVEFLRTQLD 112

Qy        113 NSVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVR 172
Db        113 NSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSMDGFALGRGLENSLPLEVVR 172
```

```
Qy        173 GAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAIT 232
              ||| |||||||||||||:||||||||||| ||||||||:||||||||||||| :||
Db        173 GAMTIRVNSLTRGHSAVRIVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAASIT 232

Qy        233 GHPDVKVHVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDS 292
              |||| ||||  :|  ||| |:|||:| ||: ||||||||||||||||||||||:| |:
Db        233 GHPDSKVHV--DG--KIMSAQEAIALKGLQPVVLGPKEGLGLVNGTAVSASMATLALTDA 288

Qy        293 HMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHE 352
              |:||||:|||||||||||||  ||| ||:||| |||| |:|||||||||| || :|| ||
Db        289 HVLSLLAQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSKYAVHHE 348

Qy        353 EEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQ 411
               |||||||||||||||||| |||:||||| ||:||:: ||||   :|||||:|:|||
Db        349 TEVKVKDDEGILRQDRYPLRCSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDLENKM 408

Qy        412 TAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYH 471
              | ||| | ||:|  :||||||:||:||::||| ||:||| ||| ||||||||||||:||
Db        409 THHGGAFMASSVGNTMEKTRLAVALMGKVSFTQLTEMLNAGMNRALPSCLAAEDPSLSYH 468

Qy        472 GKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLAS 531
               |||||  ||| ||||||||||:| |||||||||:|||||||||||||||||||||||:
Db        469 CKGLDIAAAAYTSELGHLANPVSTHVQPAEMGNQAINSLALISARRTAEANDVLSLLLAT 528

Qy        532 HLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALAL-----EVKKALNKRL 586
              |||| |||||||||| :  | |:|:: ||:|| |      |||       :|:|:: |||
Db        529 HLYCVLQAVDLRAMEFEHTKAFEPMVTELLKQHFG------ALATAEVEDKVRKSIYKRL 582

Qy        587 EQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRN 646
              :|  :|||| |||| || ||| ||| |:        |:| ::|||||| |||||:||| ||:
Db        583 QQNNSYDLEQRWHDTFSVATGAVVEALAGQ---EVSLASLNAWKVACAEKAIALTRSVRD 639

Qy        647 RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAI 706
               ||  ||| :||  ||||||||||||||||:||:|||||||::| |: |||:||||||||
Db        640 SFWAAPSSSSPALKYLSPRTRVLYSFVREEVGVKARRGDVYLGKQEVTIGTNVSRIYEAI 699

Qy        707 KDGRINHVLVKMLA 720
              | | |  |||||:|
Db        700 KSGCIAPVLVKMMA 713
```

Search completed: September 9, 2004, 10:09:38
Job time : 26.4025 secs

```
OM protein - protein search, using sw model

Run on:         September  9, 2004, 10:07:54 ; Search time 92.6141 Seconds
                                           (without alignments)
                                           2493.103 Million cell updates/sec

Title:          US-09-939-408A-13
Perfect score:  3610
Sequence:       1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters:      1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                1:  /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2:  /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3:  /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4:  /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5:  /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6:  /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7:  /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8:  /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9:  /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10:  /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11:  /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12:  /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13:  /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14:  /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15:  /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16:  /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                17:  /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                18:  /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3608	99.9	720	9	US-09-939-408A-13	Sequence 13, Appl
2	2770	76.7	726	9	US-09-939-408A-21	Sequence 21, Appl
3	2730	75.6	720	9	US-09-939-408A-30	Sequence 30, Appl
4	2614	72.4	716	9	US-09-765-873A-8	Sequence 8, Appli
5	2614	72.4	716	9	US-09-939-408A-19	Sequence 19, Appl
6	2614	72.4	716	14	US-10-138-970A-4	Sequence 4, Appli
7	2614	72.4	716	14	US-10-188-523B-8	Sequence 8, Appli
8	2614	72.4	716	15	US-10-374-366-2	Sequence 2, Appli
9	2614	72.4	716	15	US-10-439-478-2	Sequence 2, Appli
10	2613	72.4	716	9	US-09-765-873A-35	Sequence 35, Appl
11	2613	72.4	716	14	US-10-138-970A-19	Sequence 19, Appl
12	2613	72.4	716	14	US-10-188-523B-35	Sequence 35, Appl
13	2613	72.4	716	15	US-10-439-478-15	Sequence 15, Appl
14	2612	72.4	716	9	US-09-765-873A-10	Sequence 10, Appl
15	2612	72.4	716	14	US-10-138-970A-24	Sequence 24, Appl
16	2612	72.4	716	14	US-10-188-523B-10	Sequence 10, Appl
17	2612	72.4	716	15	US-10-374-366-4	Sequence 4, Appli
18	2609	72.3	716	15	US-10-374-366-24	Sequence 24, Appl
19	2608	72.2	716	15	US-10-374-366-22	Sequence 22, Appl
20	2607	72.2	716	9	US-09-765-873A-38	Sequence 38, Appl
21	2607	72.2	716	14	US-10-138-970A-22	Sequence 22, Appl
22	2607	72.2	716	14	US-10-188-523B-38	Sequence 38, Appl
23	2607	72.2	716	15	US-10-374-366-16	Sequence 16, Appl
24	2607	72.2	716	15	US-10-439-478-18	Sequence 18, Appl
25	2606	72.2	716	15	US-10-374-366-8	Sequence 8, Appli
26	2605	72.2	716	15	US-10-374-366-10	Sequence 10, Appl
27	2604	72.1	716	15	US-10-374-366-14	Sequence 14, Appl
28	2603	72.1	716	9	US-09-765-873A-36	Sequence 36, Appl
29	2603	72.1	716	9	US-09-765-873A-37	Sequence 37, Appl
30	2603	72.1	716	14	US-10-138-970A-20	Sequence 20, Appl
31	2603	72.1	716	14	US-10-138-970A-21	Sequence 21, Appl
32	2603	72.1	716	14	US-10-188-523B-36	Sequence 36, Appl
33	2603	72.1	716	14	US-10-188-523B-37	Sequence 37, Appl
34	2603	72.1	716	15	US-10-439-478-16	Sequence 16, Appl
35	2603	72.1	716	15	US-10-439-478-17	Sequence 17, Appl
36	2601	72.0	716	15	US-10-374-366-12	Sequence 12, Appl
37	2599	72.0	716	9	US-09-765-873A-34	Sequence 34, Appl
38	2599	72.0	716	14	US-10-138-970A-18	Sequence 18, Appl
39	2599	72.0	716	14	US-10-188-523B-34	Sequence 34, Appl
40	2599	72.0	716	15	US-10-374-366-18	Sequence 18, Appl
41	2599	72.0	716	15	US-10-439-478-14	Sequence 14, Appl
42	2598	72.0	716	15	US-10-374-366-6	Sequence 6, Appli
43	2598	72.0	716	15	US-10-374-366-20	Sequence 20, Appl
44	2579	71.4	716	9	US-09-765-873A-33	Sequence 33, Appl
45	2579	71.4	716	14	US-10-138-970A-17	Sequence 17, Appl

ALIGNMENTS

```
RESULT 1
US-09-939-408A-13
; Sequence 13, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Rhodotorula graminis
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (153)
;   OTHER INFORMATION: Other information:  Xaa = Val or Ala
US-09-939-408A-13

  Query Match             99.9%;  Score 3608;  DB 9;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Qy         121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Qy         181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Qy         241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Qy         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Qy         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420

Qy         421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480

Qy         481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540

Qy         541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600

Qy         601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660

Qy         661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
```

RESULT 2
US-09-939-408A-21
; Sequence 21, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)..(719)
; OTHER INFORMATION: "Xaa"means any amino acid;"Xaa"means no consensus at that posi
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOs: 13, 17, and 19

US-09-939-408A-21

```
  Query Match             76.7%;  Score 2770;  DB 9;  Length 726;
  Best Local Similarity   77.7%;  Pred. No. 5.8e-243;
  Matches  564;  Conservative   40;  Mismatches  116;  Indels    6;  Gaps    3;

Qy           1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
               ||||||:||: |||  || |||      |               :||:||:  |:|
Db           1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy          61 P-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVT 119
               | |||  || ||||: ||||||||| |||  :  ||||  ::||||:||: || ||||||
Db          61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy         120 TGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRV 179
               ||||||||||||:||||||:||||||| |||| || :|||||:||||||||| |||
Db         121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy         180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239
               |||||||||||||||||||||||| ||||||||:||||||||||||||| ||||||| ||
Db         181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy         240 HVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
               || ||| |||| |||||:||||| |||||||||||||||||||||||:|||:|||||||
Db         241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy         300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
               ||||||||||||   ||| ||:||| |||| |:||||||||||  || ||| ||||||||
Db         301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy         360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNF 418
               ||||||||||||||||:||||| ||:|::  |||     :||||:||||  | |||||
Db         361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy         419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIH 478
               |||||  :||||||||||||||||| ||:||| ||||||||||||||||:|| ||||
Db         421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy         479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
               ||| |||||||||| |||||||||||||||||||||||||||||||||||:|||| ||
Db         481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy         539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA----LALEVKKALNKRLEQTTTYDL 594
               ||||||| :||||| |     || || |                |   :| |  | |||||| :||
Db         541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy         595 EPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSS 654
               ||||||||| |||||| |||        :| |||||||| ||||||||| ||  || |||
Db         601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy         655 QAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHV 714
                :||  ||||||||||||||||:|||||||:| |: ||||||||||||| |||| |
Db         661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy         715 LVKMLA 720
               ||||||
Db         721 LVKMLA 726
```

RESULT 3
US-09-939-408A-30
; Sequence 30, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693

```
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  Preferred
;   OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
;   NAME/KEY: misc_feature
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (12)..(12)
;   OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (16)..(16)
;   OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (17)..(17)
;   OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (19)..(19)
;   OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (20)..(20)
;   OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (25)..(25)
;   OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (27)..(27)
;   OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (28)..(28)
;   OTHER INFORMATION: The 'Xaa' at location 28 stands for Ala, or Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: The 'Xaa' at location 34 stands for Thr.
;   NAME/KEY: misc_feature
;   LOCATION: (36)..(36)
;   OTHER INFORMATION: The 'Xaa' at location 36 stands for Arg, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (39)..(39)
;   OTHER INFORMATION: The 'Xaa' at location 39 stands for Ala, Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (40)..(40)
;   OTHER INFORMATION: The 'Xaa' at location 40 stands for Arg, Gly, or Trp.
;   NAME/KEY: misc_feature
;   LOCATION: (48)..(48)
;   OTHER INFORMATION: The 'Xaa' at location 48 stands for Lys, Thr, Met, Glu, Ala,   Va
;   NAME/KEY: misc_feature
;   LOCATION: (54)..(54)
;   OTHER INFORMATION: The 'Xaa' at location 54 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: The 'Xaa' at location 56 stands for Lys, Glu, or Gln.
;   NAME/KEY: misc_feature
;   LOCATION: (65)..(65)
;   OTHER INFORMATION: The 'Xaa' at location 65 stands for Glu, Asp, or Val.
;   NAME/KEY: misc_feature
;   LOCATION: (66)..(66)
;   OTHER INFORMATION: The 'Xaa' at location 66 stands for Ile, Val, or Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (76)..(76)
;   OTHER INFORMATION: The 'Xaa' at location 76 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (87)..(87)
;   OTHER INFORMATION: The 'Xaa' at location 87 stands for Thr, Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (93)..(93)
;   OTHER INFORMATION: The 'Xaa' at location 93 stands for Asp, or Ala.
;   NAME/KEY: misc_feature
;   LOCATION: (102)..(102)
;   OTHER INFORMATION: The 'Xaa' at location 102 stands for Lys, or Asn.
;   NAME/KEY: misc_feature
;   LOCATION: (103)..(103)
;   OTHER INFORMATION: The 'Xaa' at location 103 stands for Arg, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (109)..(109)
;   OTHER INFORMATION: The 'Xaa' at location 109 stands for Thr, Ala, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (112)..(112)
;   OTHER INFORMATION: The 'Xaa' at location 112 stands for Asp, His, or Tyr.
;   NAME/KEY: misc_feature
;   LOCATION: (114)..(114)
;   OTHER INFORMATION: The 'Xaa' at location 114 stands for Arg, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (117)..(117)
;   OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (148)..(148)
;   OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (150)..(150)
;   OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (153)..(153)
;   OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (154)..(154)
;   OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
;   NAME/KEY: misc_feature
;   LOCATION: (157)..(157)
;   OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
;   NAME/KEY: misc_feature
;   LOCATION: (159)..(159)
;   OTHER INFORMATION: The 'Xaa' at location 159 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (183)..(183)
;   OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
;   NAME/KEY: misc_feature
;   LOCATION: (223)..(223)
;   OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (225)..(225)
;   OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (237)..(237)
;   OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala,
;   OTHER INFORMATION: Cys, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (239)..(239)
;   OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (241)..(241)
;   OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (242)..(242)
;   OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (246)..(246)
;   OTHER INFORMATION: The 'Xaa' at location 246 stands for Lys, Asn, or Thr.
;   NAME/KEY: misc_feature
;   LOCATION: (251)..(251)
;   OTHER INFORMATION: The 'Xaa' at location 251 stands for Tyr, Ser, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (259)..(259)
;   OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (305)..(305)
;   OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
;   NAME/KEY: misc_feature
```

```
;    LOCATION: (319)..(319)
;    OTHER INFORMATION: The 'Xaa' at location 319 stands for Pro.
;    NAME/KEY: misc_feature
;    LOCATION: (346)..(346)
;    OTHER INFORMATION: The 'Xaa' at location 346 stands for Lys, Arg, Thr, Gln, Pro, a
;    OTHER INFORMATION: stop codon, Trp, or Ser.
;    NAME/KEY: misc_feature
;    LOCATION: (411)..(411)
;    OTHER INFORMATION: The 'Xaa' at location 411 stands for Lys, Thr, or Met.
;    NAME/KEY: misc_feature
;    LOCATION: (421)..(421)
;    OTHER INFORMATION: The 'Xaa' at location 421 stands for Thr, Ala, or Ser.
;    NAME/KEY: misc_feature
;    LOCATION: (457)..(457)
;    OTHER INFORMATION: The 'Xaa' at location 457 stands for Leu.
;    NAME/KEY: misc_feature
;    LOCATION: (458)..(458)
;    OTHER INFORMATION: The 'Xaa' at location 458 stands for Pro.
;    NAME/KEY: misc_feature
;    LOCATION: (466)..(466)
;    OTHER INFORMATION: The 'Xaa' at location 466 stands for Pro.
;    NAME/KEY: misc_feature
;    LOCATION: (487)..(487)
;    OTHER INFORMATION: The 'Xaa' at location 487 stands for Gly.
;    NAME/KEY: misc_feature
;    LOCATION: (493)..(493)
;    OTHER INFORMATION: The 'Xaa' at location 493 stands for Val.
;    NAME/KEY: misc_feature
;    LOCATION: (518)..(518)
;    OTHER INFORMATION: The 'Xaa' at location 518 stands for Thr.
;    NAME/KEY: misc_feature
;    LOCATION: (536)..(536)
;    OTHER INFORMATION: The 'Xaa' at location 536 stands for Val.
;    NAME/KEY: misc_feature
;    LOCATION: (556)..(556)
;    OTHER INFORMATION: The 'Xaa' at location 556 stands for Met, Ile, Val, or Leu.
;    NAME/KEY: misc_feature
;    LOCATION: (557)..(557)
;    OTHER INFORMATION: The 'Xaa' at location 557 stands for Ile, Val, or Leu.
;    NAME/KEY: misc_feature
;    LOCATION: (558)..(558)
;    OTHER INFORMATION: The 'Xaa' at location 558 stands for Thr, Ala, or Pro.

  Query Match             75.6%;  Score 2730;  DB 9;  Length 720;
  Best Local Similarity   76.5%;  Pred. No. 2.5e-239;
  Matches  551;  Conservative   45;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSXDSIATSXANGXXNXXHAAPXKXXTGATSXLXRPXXLAPPATQXTQLDIXEXILAD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         61 PTDDXXELDGYTLTLXDVVGAARKGRXVRVQTXDEIRAKIDXXVEFLRXQLXNXVYXVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        121 GFGGSADTRTEDAISLQKALLEHQLCGXLXTSXXSFXLXRGLENSLPLEVVRGAMTIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        181 SLXRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSXLXYIAAAITGHPDXKXH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        241 XXHEGXEKIMXAREAIALXGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        301 ALTAXTVEAMVGHAGSFHXFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
Db        361 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAESTTDNPLIDVENKXTHHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
Db        421 XAVANTMEKTRLALALIGKLNFTQLTEMLNAGMNRGXXSCLAAEDXSLSYHCKGLDIAAA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
Db        481 AYTSELXHLANPXTTHVQPREMGNQAVNSLALISARRXAEANDVLSLLLATHLYCXLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
Db        541 DLRAMEFEFKKQFDPXXXXLXXQHFGXALDGXELXDKVNKXLXKRLEQTNSYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
Db        601 AFSFATGTVVELLSSSXXAKVSLAAVNAWKVASAEKAISLTRXVRDXFWXAPSSSSPALX 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
Db        661 YLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGTNVSRIYEAIKXGXINHVLVKMLA 720

RESULT 4
US-09-765-873A-8
; Sequence 8, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
US-09-765-873A-8

  Query Match             72.4%;  Score 2614;  DB 9;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-229;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354
```

```
Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 5
US-09-939-408A-19
; Sequence 19, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula toruloides
US-09-939-408A-19

Query Match 72.4%; Score 2614; DB 9; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-229;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```
Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 6
US-10-138-970A-4
; Sequence 4, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: C. violaceum
US-10-138-970A-4

Query Match 72.4%; Score 2614; DB 14; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-229;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```
Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
```

```
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 7
US-10-188-523B-8
; Sequence 8, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-188-523B-8

```
  Query Match             72.4%;  Score 2614;  DB 14;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-229;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716
```

RESULT 8
US-10-374-366-2
; Sequence 2, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodosporidium glutinis
US-10-374-366-2

```
  Query Match             72.4%;  Score 2614;  DB 15;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-229;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;
```

```
Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||           |   |    |     |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| ||||  :  ||| ||||| |||::  ||||:::||||:||::|| ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||:::|| || |||||||||||||||||||||||:|||:|||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||  |||  |||:|||  ||||  |:|||  |||  ||  ||  |||  |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:|||||  |::||::  |::|    :||||||:|||||  :  ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || ||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :||||| | : :|: || |      || :: : |   :| | | |||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:   :::| |||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||::||:|||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 9
US-10-439-478-2
; Sequence 2, Application US/10439478
; Publication No. US20040018600A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sariaslani, Sima
; APPLICANT: Tang, Xao-Song
; APPLICANT: Vanelli, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: CL1912
; CURRENT APPLICATION NUMBER: US/10/439,478
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-439-478-2

```
  Query Match             72.4%;  Score 2614;  DB 15;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-229;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||           |   |    |     |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| ||||  :  ||| ||||| |||::  ||||:::||||:||::|| ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||:::|| || |||||||||||||||||||||||:|||:|||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||  |||  |||:|||  ||||  |:|||  |||  ||  ||  |||  |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:|||||  |::||::  |::|    :||||||:|||||  :  ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || ||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :||||| | : :|: || |      || :: : |   :| | | |||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:   :::| |||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||::||:|||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 10
US-09-765-873A-35
; Sequence 35, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19

```
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: X= Asp, Asn, Glu, Gln
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-35

  Query Match            72.4%;  Score 2613;  DB 9;  Length 716;
  Best Local Similarity  72.0%;  Pred. No. 1.2e-228;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716

RESULT 11
US-10-138-970A-19
; Sequence 19, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/138,970A
;  CURRENT FILING DATE:  2002-07-23
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: X=Asp, Asn, Glu, Gln
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: Xaa=Asp, Asn, Glu, Gln
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-138-970A-19

  Query Match            72.4%;  Score 2613;  DB 14;  Length 716;
  Best Local Similarity  72.0%;  Pred. No. 1.2e-228;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
```

```
               ||| |||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :||||| | : :|: || |     || :: : |  :| | | |||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:     :::| ||||||||:|| ||||||:|| ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||::||:||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 12
US-10-188-523B-35
; Sequence 35, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis mutant
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa= Asp, Asn, Glu, or Gln
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa= Thr, Ala, Ser, Pro, or Gly
US-10-188-523B-35

```
  Query Match             72.4%;  Score 2613;  DB 14;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.2e-228;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||         | | |      |       | :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| ||||  : ||| |||||| |||:: |||:::||||:||::|| |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||| |:|   || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:|||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||  ||| ||:||| |||| |:|||  ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:||||| |::||:: |::|   :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||||  || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| |||||| ||||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :||||| | : :|: || |     || :: : |  :| | | |||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:     :::| ||||||||:|| ||||||:|| ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||::||:||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 13
US-10-439-478-15
; Sequence 15, Application US/10439478
; Publication No. US20040018600A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sariaslani, Sima
; APPLICANT: Tang, Xao-Song
; APPLICANT: Vanelli, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: CL1912
; CURRENT APPLICATION NUMBER: US/10/439,478
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
; LENGTH: 716
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa= Asp, Asn, Glu, or Gln
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa= Thr, Ala, Ser, Pro, or Gly
US-10-439-478-15

```
  Query Match             72.4%;  Score 2613;  DB 15;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.2e-228;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;
```

```
Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |      |         |         :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| ||||  : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||||:||||||:|||||||  |:|    ||  :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| ||||||||:|||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:|||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||   ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:||||| |::||:: |::|    :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||||  ||  ||||||| ||:|||  |||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| ||||||||||||||  |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||  | :|||||  | : :|: || |    ||  :: :  |    :|  | | ||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| ||||:|||:     :::| ||||||||:|| ||||||:||  ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 14
US-09-765-873A-10
; Sequence 10, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis

US-09-765-873A-10

```
  Query Match             72.4%;  Score 2612;  DB 9;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.4e-228;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |      |         |         :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| ||||  : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||||:||||||:|||||||  |:|    ||  :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| ||||||||:|||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:|||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||   ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:||||| |::||:: |::|    :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||||  ||  ||||||| ||:|||  |||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| ||||||||||||||  |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||  | :|||||  | : :|: || |    ||  :: :  |    :|  | | ||||| :|||
Db        535 IDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| ||||:|||:     :::| ||||||||:|| ||||||:||  ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 15
US-10-138-970A-24
; Sequence 24, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24

```
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: artificial sequence
;   FEATURE:
;   OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-138-970A-24

  Query Match             72.4%;  Score 2612;  DB 14;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.4e-228;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||       |  | |    |      |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||||:||||||:||||||| |:|  || :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| |||||||||:|||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||  ||| ||:||| |||| |:||| ||| || || ||| ||||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:||||| |::|| : |::|   :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| ||||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :||||| | : :|: || |   ||  : : |  :| | | | |||||| :|||
Db        535 IDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:   :::| |||||||| || ||||||:||: ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716

Search completed: September  9, 2004, 10:24:24
Job time : 95.6141 secs
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         September  9, 2004, 10:04:14 ; Search time 295.768 Seconds
                                           (without alignments)
                                           2376.050 Million cell updates/sec

Title:          US-09-939-408A-13
Perfect score:  3610
Sequence:       1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters:      6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Pending_Patents_AA_Main:*
                1:  /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
                2:  /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
                3:  /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
                4:  /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
                5:  /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
                6:  /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
                7:  /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
                8:  /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
                9:  /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
                10:  /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
                11:  /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
                12:  /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
                13:  /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
                14:  /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
                15:  /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
                16:  /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
                17:  /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
                18:  /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
                19:  /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
                20:  /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
                21:  /cgn2_6/ptodata/2/paa/US097A_COMB.pep:*
                22:  /cgn2_6/ptodata/2/paa/US097B_COMB.pep:*
                23:  /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
                24:  /cgn2_6/ptodata/2/paa/US099A_COMB.pep:*
                25:  /cgn2_6/ptodata/2/paa/US099B_COMB.pep:*
                26:  /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
                27:  /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
                28:  /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
                29:  /cgn2_6/ptodata/2/paa/US103_COMB.pep:*
                30:  /cgn2_6/ptodata/2/paa/US104_COMB.pep:*
                31:  /cgn2_6/ptodata/2/paa/US106_COMB.pep:*
                32:  /cgn2_6/ptodata/2/paa/US107_COMB.pep:*
                33:  /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                %
Result          Query
  No.   Score   Match Length DB  ID                               Description
---------------------------------------------------------------------------
    1    3608    99.9    720  1  PCT-US02-26934A-13          Sequence 13, Appl
    2    3608    99.9    720 24  US-09-939-408A-13           Sequence 13, Appl
    3    2770    76.7    726  1  PCT-US02-26934A-21          Sequence 21, Appl
    4    2770    76.7    726 24  US-09-939-408A-21           Sequence 21, Appl
    5    2730    75.6    720  1  PCT-US02-26934A-30          Sequence 30, Appl
    6    2730    75.6    720 24  US-09-939-408A-30           Sequence 30, Appl
    7    2614    72.4    716  1  PCT-US02-26934A-19          Sequence 19, Appl
    8    2614    72.4    716  1  PCT-US03-05708-2            Sequence 2, Appli
    9    2614    72.4    716  1  PCT-US03-17926-2            Sequence 2, Appli
   10    2614    72.4    716 24  US-09-939-408A-19           Sequence 19, Appl
   11    2614    72.4    716 27  US-10-138-970A-4            Sequence 4, Appli
   12    2614    72.4    716 27  US-10-188-523B-8            Sequence 8, Appli
   13    2614    72.4    716 27  US-10-188-523C-8            Sequence 8, Appli
   14    2614    72.4    716 29  US-10-374-366-2             Sequence 2, Appli
   15    2614    72.4    716 30  US-10-439-478-2             Sequence 2, Appli
   16    2614    72.4    716 33  US-60-147-719-8             Sequence 8, Appli
   17    2614    72.4    716 33  US-60-383-089-4             Sequence 4, Appli
   18    2614    72.4    716 33  US-60-460-120-4             Sequence 4, Appli
   19    2613    72.4    716  1  PCT-US03-17926-15           Sequence 15, Appl
   20    2613    72.4    716 27  US-10-138-970A-19           Sequence 19, Appl
   21    2613    72.4    716 27  US-10-188-523B-35           Sequence 35, Appl
   22    2613    72.4    716 27  US-10-188-523C-35           Sequence 35, Appl
   23    2613    72.4    716 30  US-10-439-478-15            Sequence 15, Appl
   24    2612    72.4    716  1  PCT-US03-05708-4            Sequence 4, Appli
   25    2612    72.4    716 27  US-10-138-970A-24           Sequence 24, Appl
   26    2612    72.4    716 27  US-10-188-523B-10           Sequence 10, Appl
   27    2612    72.4    716 27  US-10-188-523C-10           Sequence 10, Appl
   28    2612    72.4    716 29  US-10-374-366-4             Sequence 4, Appli
   29    2612    72.4    716 33  US-60-147-719-10            Sequence 10, Appl
   30    2609    72.3    716  1  PCT-US03-05708-24           Sequence 24, Appl
   31    2609    72.3    716 29  US-10-374-366-24            Sequence 24, Appl
   32    2608    72.2    716  1  PCT-US03-05708-22           Sequence 22, Appl
   33    2608    72.2    716 29  US-10-374-366-22            Sequence 22, Appl
   34    2607    72.2    716  1  PCT-US03-05708-16           Sequence 16, Appl
   35    2607    72.2    716  1  PCT-US03-17926-18           Sequence 18, Appl
   36    2607    72.2    716 27  US-10-138-970A-22           Sequence 22, Appl
   37    2607    72.2    716 27  US-10-188-523B-38           Sequence 38, Appl
   38    2607    72.2    716 27  US-10-188-523C-38           Sequence 38, Appl
   39    2607    72.2    716 29  US-10-374-366-16            Sequence 16, Appl
   40    2607    72.2    716 30  US-10-439-478-18            Sequence 18, Appl
   41    2606    72.2    716  1  PCT-US03-05708-8            Sequence 8, Appli
   42    2606    72.2    716 29  US-10-374-366-8             Sequence 8, Appli
   43    2605    72.2    716  1  PCT-US03-05708-10           Sequence 10, Appl
   44    2605    72.2    716 29  US-10-374-366-10            Sequence 10, Appl
   45    2604    72.1    716  1  PCT-US03-05708-14           Sequence 14, Appl

                                  ALIGNMENTS

RESULT 1
PCT-US02-26934A-13
; Sequence 13, Application PC/TUS0226934A
; GENERAL INFORMATION:
;  APPLICANT: PCBU Services, Inc. et al.
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA PCT
;  CURRENT APPLICATION NUMBER: PCT/US02/26934A
;  CURRENT FILING DATE:  2002-08-23
;  PRIOR APPLICATION NUMBER: US 09/939,408
;  PRIOR FILING DATE: 2001-08-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Rhodotorula graminis
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (153)
;   OTHER INFORMATION: Other information:  Xaa = Val or Ala
```

```
PCT-US02-26934A-13

  Query Match             99.9%;  Score 3608;  DB 1;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720


RESULT 2
US-09-939-408A-13
; Sequence 13, Application US/09939408A
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Rhodotorula graminis
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (153)
;   OTHER INFORMATION: Other information:  Xaa = Val or Ala
US-09-939-408A-13

  Query Match             99.9%;  Score 3608;  DB 24;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  720;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720


RESULT 3
PCT-US02-26934A-21
; Sequence 21, Application PC/TUS0226934A
; GENERAL INFORMATION:
;  APPLICANT: PCBU Services, Inc. et al.
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA PCT
;  CURRENT APPLICATION NUMBER: PCT/US02/26934A
```

```
;  CURRENT FILING DATE:  2002-08-23
;  PRIOR APPLICATION NUMBER: US 09/939,408
;  PRIOR FILING DATE: 2001-08-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (12)..(719)
;   OTHER INFORMATION: "Xaa"means any amino acid;"Xaa"means no consensus at that positic
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus of
;   OTHER INFORMATION: SEQ ID NOs: 13, 17, and 19
PCT-US02-26934A-21

  Query Match             76.7%;  Score 2770;  DB 1;  Length 726;
  Best Local Similarity   77.7%;  Pred. No. 4.3e-262;
  Matches  564;  Conservative   40;  Mismatches  116;  Indels    6;  Gaps    3;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||:||: |||  || |||    |            |         :||:||:  |:|
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy         61 P-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVT 119
              | |||  || |||||: ||||||||| |||  : ||||  ::||||:||: || |||||||
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        120 TGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRV 179
              |||||||||||||:||||||:|||||| |||| || :||||||:||||||||||| |||
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239
              ||||||||||||||||||||||| |||||||||:|||||||||||||| ||||||| ||
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        240 HVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
              || ||| |||| |||||:||||| ||||||||||||||||||||||||:|||:|||||||
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
              |||||||||||| ||| ||:||| |||| |:||||||||| || ||| ||||||||||
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNF 418
              ||||||||||||||||:|||||  ||:||:: ||||   :||||||:||||| | |||||
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIH 478
              |||||  :|||||||||||||||||||| ||:||| ||||||||||||||||:|| |||||
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
               ||| |||||||||||| |||||||||||||||||||||||||||||||||||:|||| ||
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA----LALEVKKALNKRLEQTTTYDL 594
              ||||||| :||||| |      || || |             | :| | | |||||   :|||
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy        595 EPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSS 654
              |||||||| |||||| |||      |:| |||||||| ||||||||| || ||  || |||
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        655 QAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHV 714
               :||  ||||||||||||||||||:|||||||:| |: |||||||||||||| |||| |
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        715 LVKMLA 720
              ||||||
Db        721 LVKMLA 726
```

```
RESULT 4
US-09-939-408A-21
; Sequence 21, Application US/09939408A
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (12)..(719)
;   OTHER INFORMATION: "Xaa"means any amino acid;"Xaa"means no consensus at that    posi
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus of
;   OTHER INFORMATION: SEQ ID NOs: 13, 17, and 19
US-09-939-408A-21

  Query Match             76.7%;  Score 2770;  DB 24;  Length 726;
  Best Local Similarity   77.7%;  Pred. No. 4.3e-262;
  Matches  564;  Conservative   40;  Mismatches  116;  Indels    6;  Gaps    3;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||:||: |||  || |||    |            |         :||:||:  |:|
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy         61 P-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVT 119
              | |||  || |||||: ||||||||| |||  : ||||  ::||||:||: || |||||||
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        120 TGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRV 179
              |||||||||||||:||||||:|||||| |||| || :||||||:||||||||||| |||
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239
              ||||||||||||||||||||||| |||||||||:|||||||||||||| ||||||| ||
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        240 HVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
              || ||| |||| |||||:||||| ||||||||||||||||||||||||:|||:|||||||
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
              |||||||||||| ||| ||:||| |||| |:||||||||| || ||| ||||||||||
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNF 418
              ||||||||||||||||:|||||  ||:||:: ||||   :||||||:||||| | |||||
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIH 478
              |||||  :|||||||||||||||||||| ||:||| ||||||||||||||||:|| |||||
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
               ||| |||||||||||| |||||||||||||||||||||||||||||||||||:|||| ||
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA----LALEVKKALNKRLEQTTTYDL 594
```

```
              ||||||| :||||| |     || || |              |  :| | | |||||| :|||
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy        595 EPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSS 654
              |||||||| |||||| |||      |:| ||||||| |||||||| ||  || |||
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        655 QAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHV 714
                :||  ||||||||||||||||||:|||||||:| |: |||||||||||| |||| |
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        715 LVKMLA 720
              ||||||
Db        721 LVKMLA 726
```

RESULT 5
PCT-US02-26934A-30
; Sequence 30, Application PC/TUS0226934A
; GENERAL INFORMATION:
; APPLICANT: PCBU Services, Inc. et al.
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA PCT
; CURRENT APPLICATION NUMBER: PCT/US02/26934A
; CURRENT FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: US 09/939,408
; PRIOR FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Preferred
; OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5)..(5)
; OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12)..(12)
; OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (16)..(16)
; OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (17)..(17)
; OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (19)..(19)
; OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (25)..(25)
; OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (27)..(27)
; OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28)..(28)
; OTHER INFORMATION: The 'Xaa' at location 28 stands for Ala, or Pro.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34)..(34)
; OTHER INFORMATION: The 'Xaa' at location 34 stands for Thr.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36)..(36)
; OTHER INFORMATION: The 'Xaa' at location 36 stands for Arg, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (39)..(39)
; OTHER INFORMATION: The 'Xaa' at location 39 stands for Ala, Pro, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40)..(40)
; OTHER INFORMATION: The 'Xaa' at location 40 stands for Arg, Gly, or Trp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (48)..(48)
; OTHER INFORMATION: The 'Xaa' at location 48 stands for Lys, Thr, Met, Glu, Ala,
; OTHER INFORMATION: Val, Gln, Pro, or Leu.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (54)..(54)
; OTHER INFORMATION: The 'Xaa' at location 54 stands for Val.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (56)..(56)
; OTHER INFORMATION: The 'Xaa' at location 56 stands for Lys, Glu, or Gln.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65)..(65)
; OTHER INFORMATION: The 'Xaa' at location 65 stands for Glu, Asp, or Val.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (66)..(66)
; OTHER INFORMATION: The 'Xaa' at location 66 stands for Ile, Val, or Leu.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (76)..(76)
; OTHER INFORMATION: The 'Xaa' at location 76 stands for Gly.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (87)..(87)
; OTHER INFORMATION: The 'Xaa' at location 87 stands for Thr, Pro, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (93)..(93)
; OTHER INFORMATION: The 'Xaa' at location 93 stands for Asp, or Ala.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102)..(102)
; OTHER INFORMATION: The 'Xaa' at location 102 stands for Lys, or Asn.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103)..(103)
; OTHER INFORMATION: The 'Xaa' at location 103 stands for Arg, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (109)..(109)
; OTHER INFORMATION: The 'Xaa' at location 109 stands for Thr, Ala, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (112)..(112)
; OTHER INFORMATION: The 'Xaa' at location 112 stands for Asp, His, or Tyr.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (114)..(114)
; OTHER INFORMATION: The 'Xaa' at location 114 stands for Arg, or Ser.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117)..(117)

```
;   OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (148)..(148)
;   OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150)..(150)
;   OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (153)..(153)
;   OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (154)..(154)
;   OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (157)..(157)
;   OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (159)..(159)
;   OTHER INFORMATION: The 'Xaa' at location 159 stands for Gly.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (183)..(183)
;   OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (223)..(223)
;   OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (225)..(225)
;   OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (237)..(237)
;   OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala, Val
;   OTHER INFORMATION: Cys, or Phe.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (239)..(239)
;   OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (241)..(241)
;   OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (242)..(242)
;   OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (246)..(246)
;   OTHER INFORMATION: The 'Xaa' at location 246 stands for Lys, Asn, or Thr.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (251)..(251)
;   OTHER INFORMATION: The 'Xaa' at location 251 stands for Tyr, Ser, or Phe.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (259)..(259)
;   OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (305)..(305)
;   OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
;   FEATURE:
;   NAME/KEY: misc_feature

  Query Match             75.6%;  Score 2730;  DB 1;  Length 720;
  Best Local Similarity   76.5%;  Pred. No. 3.7e-258;
  Matches  551;  Conservative   45;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSXDSIATSXANGXXNXXHAAPXKXXTGATSXLXRPXXLAPPATQXTQLDIXEXILAD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         61 PTDDXXELDGYTLTLXDVVGAARKGRXVRVQTXDEIRAKIDXXVEFLRXQLXNXVYXVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        121 GFGGSADTRTEDAISLQKALLEHQLCGXLXTSXXSFXLXRGLENSLPLEVVRGAMTIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        181 SLXRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSXLXYIAAAITGHPDXKXH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        241 XXHEGXEKIMXAREAIALXGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        301 ALTAXTVEAMVGHAGSFHXFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
Db        361 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAESTTDNPLIDVENKXTHHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
Db        421 XAVANTMEKTRLALALIGKLNFTQLTEMLNAGMNRGXXSCLAAEDXSLSYHCKGLDIAAA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
Db        481 AYTSELXHLANPXTTHVQPREMGNQAVNSLALISARRXAEANDVLSLLLATHLYCXLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
Db        541 DLRAMEFEFKKQFDPXXXXLXXQHFGXALDGXELXDKVNKXLXKRLEQTNSYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
Db        601 AFSFATGTVVELLSSSXXAKVSLAAVNAWKVASAEKAISLTRXVRDXFWXAPSSSSPALX 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
Db        661 YLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGTNVSRIYEAIKXGXINHVLVKMLA 720

RESULT 6
US-09-939-408A-30
; Sequence 30, Application US/09939408A
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
;   LENGTH: 720
```

```
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   Preferred
;   OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
;   NAME/KEY: misc_feature
;   LOCATION: (5)..(5)
;   OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (12)..(12)
;   OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (16)..(16)
;   OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (17)..(17)
;   OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (19)..(19)
;   OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (20)..(20)
;   OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (25)..(25)
;   OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (27)..(27)
;   OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (28)..(28)
;   OTHER INFORMATION: The 'Xaa' at location 28 stands for Ala, or Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (34)..(34)
;   OTHER INFORMATION: The 'Xaa' at location 34 stands for Thr.
;   NAME/KEY: misc_feature
;   LOCATION: (36)..(36)
;   OTHER INFORMATION: The 'Xaa' at location 36 stands for Arg, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (39)..(39)
;   OTHER INFORMATION: The 'Xaa' at location 39 stands for Ala, Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (40)..(40)
;   OTHER INFORMATION: The 'Xaa' at location 40 stands for Arg, Gly, or Trp.
;   NAME/KEY: misc_feature
;   LOCATION: (48)..(48)
;   OTHER INFORMATION: The 'Xaa' at location 48 stands for Lys, Thr, Met, Glu, Ala,   Va
;   NAME/KEY: misc_feature
;   LOCATION: (54)..(54)
;   OTHER INFORMATION: The 'Xaa' at location 54 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (56)..(56)
;   OTHER INFORMATION: The 'Xaa' at location 56 stands for Lys, Glu, or Gln.
;   NAME/KEY: misc_feature
;   LOCATION: (65)..(65)
;   OTHER INFORMATION: The 'Xaa' at location 65 stands for Glu, Asp, or Val.
;   NAME/KEY: misc_feature
;   LOCATION: (66)..(66)
;   OTHER INFORMATION: The 'Xaa' at location 66 stands for Ile, Val, or Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (76)..(76)
;   OTHER INFORMATION: The 'Xaa' at location 76 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (87)..(87)
;   OTHER INFORMATION: The 'Xaa' at location 87 stands for Thr, Pro, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (93)..(93)
;   OTHER INFORMATION: The 'Xaa' at location 93 stands for Asp, or Ala.
;   NAME/KEY: misc_feature
;   LOCATION: (102)..(102)
;   OTHER INFORMATION: The 'Xaa' at location 102 stands for Lys, or Asn.
;   NAME/KEY: misc_feature
;   LOCATION: (103)..(103)
;   OTHER INFORMATION: The 'Xaa' at location 103 stands for Arg, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (109)..(109)
;   OTHER INFORMATION: The 'Xaa' at location 109 stands for Thr, Ala, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (112)..(112)
;   OTHER INFORMATION: The 'Xaa' at location 112 stands for Asp, His, or Tyr.
;   NAME/KEY: misc_feature
;   LOCATION: (114)..(114)
;   OTHER INFORMATION: The 'Xaa' at location 114 stands for Arg, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (117)..(117)
;   OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (148)..(148)
;   OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (150)..(150)
;   OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (153)..(153)
;   OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (154)..(154)
;   OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
;   NAME/KEY: misc_feature
;   LOCATION: (157)..(157)
;   OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
;   NAME/KEY: misc_feature
;   LOCATION: (159)..(159)
;   OTHER INFORMATION: The 'Xaa' at location 159 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (183)..(183)
;   OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
;   NAME/KEY: misc_feature
;   LOCATION: (223)..(223)
;   OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (225)..(225)
;   OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (237)..(237)
;   OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala,
;   OTHER INFORMATION: Cys, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (239)..(239)
;   OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (241)..(241)
;   OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (242)..(242)
;   OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (246)..(246)
;   OTHER INFORMATION: The 'Xaa' at location 246 stands for Lys, Asn, or Thr.
;   NAME/KEY: misc_feature
;   LOCATION: (251)..(251)
;   OTHER INFORMATION: The 'Xaa' at location 251 stands for Tyr, Ser, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (259)..(259)
;   OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
;   NAME/KEY: misc_feature
;   LOCATION: (305)..(305)
;   OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (319)..(319)
;   OTHER INFORMATION: The 'Xaa' at location 319 stands for Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (346)..(346)
;   OTHER INFORMATION: The 'Xaa' at location 346 stands for Lys, Arg, Thr, Gln, Pro, a
;   OTHER INFORMATION: stop codon, Trp, or Ser.
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (411)..(411)
;   OTHER INFORMATION: The 'Xaa' at location 411 stands for Lys, Thr, or Met.
;   NAME/KEY: misc_feature
;   LOCATION: (421)..(421)
;   OTHER INFORMATION: The 'Xaa' at location 421 stands for Thr, Ala, or Ser.
;   NAME/KEY: misc_feature
;   LOCATION: (457)..(457)
;   OTHER INFORMATION: The 'Xaa' at location 457 stands for Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (458)..(458)
;   OTHER INFORMATION: The 'Xaa' at location 458 stands for Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (466)..(466)
;   OTHER INFORMATION: The 'Xaa' at location 466 stands for Pro.
;   NAME/KEY: misc_feature
;   LOCATION: (487)..(487)
;   OTHER INFORMATION: The 'Xaa' at location 487 stands for Gly.
;   NAME/KEY: misc_feature
;   LOCATION: (493)..(493)
;   OTHER INFORMATION: The 'Xaa' at location 493 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (518)..(518)
;   OTHER INFORMATION: The 'Xaa' at location 518 stands for Thr.
;   NAME/KEY: misc_feature
;   LOCATION: (536)..(536)
;   OTHER INFORMATION: The 'Xaa' at location 536 stands for Val.
;   NAME/KEY: misc_feature
;   LOCATION: (556)..(556)
;   OTHER INFORMATION: The 'Xaa' at location 556 stands for Met, Ile, Val, or Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (557)..(557)
;   OTHER INFORMATION: The 'Xaa' at location 557 stands for Ile, Val, or Leu.
;   NAME/KEY: misc_feature
;   LOCATION: (558)..(558)
;   OTHER INFORMATION: The 'Xaa' at location 558 stands for Thr, Ala, or Pro.
;   NAME/KEY: misc_feature

  Query Match             75.6%;  Score 2730;  DB 24;  Length 720;
  Best Local Similarity   76.5%;  Pred. No. 3.7e-258;
  Matches  551;  Conservative   45;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSXDSIATSXANGXXNXXHAAPXKXXTGATSXLXRPXXLAPPATQXTQLDIXEXILAD 60

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         61 PTDDXXELDGYTLTLXDVVGAARKGRXVRVQTXDEIRAKIDXXVEFLRXQLXNXVYXVTT 120

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        121 GFGGSADTRTEDAISLQKALLEHQLCGXLXTSXXSFXLXRGLENSLPLEVVRGAMTIRVN 180

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        181 SLXRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSXLXYIAAAITGHPDXKXH 240

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        241 XXHEGXEKIMXAREAIALXGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 300

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        301 ALTAXTVEAMVGHAGSFHXFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDD 360

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQA 420
Db        361 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAESTTDNPLIDVENKXTHHGGNFQA 420

Qy        421 SAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHIA 480
Db        421 XAVANTMEKTRLALALIGKLNFTQLTEMLNAGMNRGXXSCLAAEDXSLSYHCKGLDIAAA 480

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
Db        481 AYTSELXHLANPXTTHVQPREMGNQAVNSLALISARRXAEANDVLSLLLATHLYCXLQAV 540

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHD 600
Db        541 DLRAMEFEFKKQFDPXXXXLXXQHFGXALDGXELXDKVNKXLXKRLEQTNSYDLEPRWHD 600

Qy        601 AFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHA 660
Db        601 AFSFATGTVVELLSSSXXAKVSLAAVNAWKVASAEKAISLTRXVRDXFWXAPSSSSPALX 660

Qy        661 YLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA 720
Db        661 YLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGTNVSRIYEAIKXGXINHVLVKMLA 720

RESULT 7
PCT-US02-26934A-19
; Sequence 19, Application PC/TUS0226934A
; GENERAL INFORMATION:
;  APPLICANT: PCBU Services, Inc. et al.
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA PCT
;  CURRENT APPLICATION NUMBER: PCT/US02/26934A
;  CURRENT FILING DATE:  2002-08-23
;  PRIOR APPLICATION NUMBER: US 09/939,408
;  PRIOR FILING DATE: 2001-08-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula toruloides
PCT-US02-26934A-19

  Query Match             72.4%;  Score 2614;  DB 1;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474
```

```
Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| ||||| |||||||||||||| |:||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :||||| | : :|: || |      || :: : |   :| | | ||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:|| ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:|||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 8
PCT-US03-05708-2
; Sequence 2, Application PC/TUS0305708
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours and Company, Inc.
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/05708
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodosporidium glutinis
PCT-US03-05708-2

```
  Query Match             72.4%;  Score 2614;  DB 1;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||         | | |     |           :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| ||||||||:|||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || |||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||  ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:||||| |::|::  |::|    :|||||:|||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414
```

```
Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || |||||||| ||:||| |||||||||||||||||:|| ||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| ||||| |||||||||||||| |:||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :||||| | : :|: || |      || :: : |   :| | | ||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:|| ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:|||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 9
PCT-US03-17926-2
; Sequence 2, Application PC/TUS0317926
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: CL1912 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/17926
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
PCT-US03-17926-2

```
  Query Match             72.4%;  Score 2614;  DB 1;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||         | | |     |           :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| ||||||||:|||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || |||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||  ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
```

```
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479

Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539

Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595

Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655

Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715

Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720

Db        712 LKMLA 716


RESULT 10
US-09-939-408A-19
; Sequence 19, Application US/09939408A
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula toruloides
US-09-939-408A-19

  Query Match             72.4%;  Score 2614;  DB 24;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419

Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479

Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539

Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595

Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655

Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715

Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720

Db        712 LKMLA 716


RESULT 11
US-10-138-970A-4
; Sequence 4, Application US/10138970A
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/138,970A
;  CURRENT FILING DATE:  2002-07-23
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: C. violaceum
US-10-138-970A-4

  Query Match             72.4%;  Score 2614;  DB 27;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
```

```
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419

Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479

Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539

Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595

Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655

Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715

Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720

Db        712 LKMLA 716
```

```
RESULT 12
US-10-188-523B-8
; Sequence 8, Application US/10188523B
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US DIVCIP
;  CURRENT APPLICATION NUMBER: US/10/188,523B
;  CURRENT FILING DATE:  2002-10-17
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
US-10-188-523B-8

  Query Match             72.4%;  Score 2614;  DB 27;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114
```

```
Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419

Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479

Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539

Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595

Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655

Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715

Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720

Db        712 LKMLA 716
```

```
RESULT 13
US-10-188-523C-8
; Sequence 8, Application US/10188523C
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009-CIP
;  CURRENT APPLICATION NUMBER: US/10/188,523C
;  CURRENT FILING DATE:  2002-10-24
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 8
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
US-10-188-523C-8

  Query Match             72.4%;  Score 2614;  DB 27;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54
```

```
Qy          61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
               |||   :|| ||||  :  |||  ||||||  |||::  ||||:::||||:||::||  ||||||
Db          55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy         121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
               ||||||||||||:||||||:|||||||  |:|    ||  :|||||||:||||||||||  ||||
Db         115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy         181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
               ||||||||||||||||||||||||  |||||||||:||||||||||||||||  ||:||||  |||
Db         175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy         241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
               |:|||  |||::||||::||  ||  ||||||||||||||||||||||||||||:|||:||||||||
Db         235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
               :|||:||||||||   |||  |||:|||  ||||  |:|||  |||  ||  ||  |||  |||||||||
Db         295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
               ||||||||||||||||:|||||  |::||:: |::|    :|||||:||||| :  ||||||
Db         355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy         420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
               |:||: :||||||  ||  ||||||||  ||:|||  ||||||||||||||||:||  |||||
Db         415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy         480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
               |||  ||||||||||||  ||||||  |||||||||||||||  |:|||||||||:||||  |||
Db         475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy         540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
               :||||:|  :|||||  |  :  :|:  ||  |     ||  ::  :  |    :|  |  |  ||||||  :|||
Db         535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy         596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
               ||||||||:|  |||||:|||:     :::|  ||||||||:||  |||||||:||  ||    |:
Db         595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy         656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
               :||  :||||||::||:||||||||:|||||||:|  |:  ||||||:||||||  ||||:||
Db         652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy         716 VKMLA 720
               :||||
Db         712 LKMLA 716
```

RESULT 14
US-10-374-366-2
; Sequence 2, Application US/10374366
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-2

```
  Query Match             72.4%;  Score 2614;  DB 29;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy           1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
               |||||||::  :  |||          |   |   |       |        |          :|::||:::|:
Db           1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy          61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
               |||   :|| ||||  :  |||  ||||||  |||::  ||||:::||||:||::||  ||||||
Db          55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy         121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
               ||||||||||||:||||||:|||||||  |:|    ||  :|||||||:||||||||||  ||||
Db         115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy         181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
               ||||||||||||||||||||||||  |||||||||:||||||||||||||||  ||:||||  |||
Db         175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy         241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
               |:|||  |||::||||::||  ||  ||||||||||||||||||||||||||||:|||:||||||||
Db         235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
               :|||:||||||||   |||  |||:|||  ||||  |:|||  |||  ||  ||  |||  |||||||||
Db         295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
               ||||||||||||||||:|||||  |::||:: |::|    :|||||:||||| :  ||||||
Db         355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy         420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
               |:||: :||||||  ||  ||||||||  ||:|||  ||||||||||||||||:||  |||||
Db         415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy         480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
               |||  ||||||||||||  ||||||  |||||||||||||||  |:|||||||||:||||  |||
Db         475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy         540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
               :||||:|  :|||||  |  :  :|:  ||  |     ||  ::  :  |    :|  |  |  ||||||  :|||
Db         535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy         596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
               ||||||||:|  |||||:|||:     :::|  ||||||||:||  |||||||:||  ||    |:
Db         595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy         656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
               :||  :||||||::||:||||||||:|||||||:|  |:  ||||||:||||||  ||||:||
Db         652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy         716 VKMLA 720
               :||||
Db         712 LKMLA 716
```

RESULT 15
US-10-439-478-2
; Sequence 2, Application US/10439478
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sariaslani, Sima
; APPLICANT: Tang, Xao-Song
; APPLICANT: Vanelli, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: CL1912
; CURRENT APPLICATION NUMBER: US/10/439,478
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2

```
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
US-10-439-478-2

  Query Match             72.4%;  Score 2614;  DB 30;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 9.5e-247;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716

Search completed: September  9, 2004, 10:19:37
Job time : 298.768 secs
```

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:04:49 ; Search time 43.8174 Seconds
(without alignments)
2131.280 Million cell updates/sec

Title: US-09-939-408A-13
Perfect score: 3610
Sequence: 1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 704768 seqs, 129704438 residues

Total number of hits satisfying chosen parameters: 704768

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2614	72.4	716	6	US-10-476-193-4	Sequence 4, Appli
2	2614	72.4	716	6	US-10-476-198-4	Sequence 4, Appli
3	2613	72.4	716	6	US-10-476-193-19	Sequence 19, Appl
4	2613	72.4	716	6	US-10-476-198-19	Sequence 19, Appl
5	2612	72.4	716	6	US-10-476-193-24	Sequence 24, Appl
6	2612	72.4	716	6	US-10-476-198-24	Sequence 24, Appl
7	2607	72.2	716	6	US-10-476-193-22	Sequence 22, Appl
8	2607	72.2	716	6	US-10-476-198-22	Sequence 22, Appl
9	2603	72.1	716	6	US-10-476-193-20	Sequence 20, Appl
10	2603	72.1	716	6	US-10-476-193-21	Sequence 21, Appl
11	2603	72.1	716	6	US-10-476-198-20	Sequence 20, Appl
12	2603	72.1	716	6	US-10-476-198-21	Sequence 21, Appl
13	2599	72.0	716	6	US-10-476-193-18	Sequence 18, Appl
14	2599	72.0	716	6	US-10-476-198-18	Sequence 18, Appl
15	2579	71.4	716	6	US-10-476-193-17	Sequence 17, Appl
16	2579	71.4	716	6	US-10-476-198-17	Sequence 17, Appl
17	1036	28.7	712	6	US-10-491-823-332	Sequence 332, App
18	1031	28.6	712	6	US-10-491-823-334	Sequence 334, App
19	1010.5	28.0	714	6	US-10-769-350-10	Sequence 10, Appl
20	1005	27.8	703	6	US-10-425-115-267547	Sequence 267547,
21	1003.5	27.8	671	7	US-60-548-891-28	Sequence 28, Appl
22	1002.5	27.8	703	6	US-10-425-115-339651	Sequence 339651,
23	999	27.7	716	6	US-10-425-115-218409	Sequence 218409,
24	996.5	27.6	705	6	US-10-767-701-47379	Sequence 47379, A
25	993	27.5	714	6	US-10-425-115-218412	Sequence 218412,
26	990	27.4	668	6	US-10-767-701-47377	Sequence 47377, A
27	985.5	27.3	713	7	US-60-563-879-184	Sequence 184, App
28	985	27.3	717	6	US-10-425-115-217555	Sequence 217555,
29	983.5	27.2	704	6	US-10-425-115-218411	Sequence 218411,
30	983	27.2	711	7	US-60-563-879-108	Sequence 108, App
31	982	27.2	711	6	US-10-769-350-13	Sequence 13, Appl
32	980.5	27.2	713	7	US-60-563-879-109	Sequence 109, App
33	979	27.1	706	6	US-10-425-115-339653	Sequence 339653,
34	973.5	27.0	722	6	US-10-425-115-217553	Sequence 217553,
35	971.5	26.9	672	6	US-10-425-115-339642	Sequence 339642,
36	967	26.8	671	6	US-10-491-823-330	Sequence 330, App
37	966.5	26.8	698	6	US-10-425-115-327033	Sequence 327033,
38	892.5	24.7	619	6	US-10-425-115-217557	Sequence 217557,
39	865.5	24.0	617	6	US-10-425-115-267544	Sequence 267544,
40	854.5	23.7	584	6	US-10-425-115-267542	Sequence 267542,
41	565.5	15.7	499	6	US-10-159-257A-189	Sequence 189, App
42	506.5	14.0	510	7	US-60-579-902-6176	Sequence 6176, Ap
43	494.5	13.7	277	6	US-10-425-115-267539	Sequence 267539,
44	448.5	12.4	192	6	US-10-491-823-269	Sequence 269, App
45	434.5	12.0	281	6	US-10-425-115-352913	Sequence 352913,

ALIGNMENTS

RESULT 1
US-10-476-193-4
; Sequence 4, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: C. violaceum
US-10-476-193-4

Query Match 72.4%; Score 2614; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 3.2e-194;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```
Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||      | | |     |             :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| ||||  : ||| |||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||||  |:|   || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||||| ||||||||:||||||||||||||| ||:||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:|||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||   ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
```

```
             ||||||||||||||||:||||| |::||:: |::|     :||||||:||||| : ||||||
Db       355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy       420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
             |:||: :|||||| || ||||||| ||:||| |||||||||||||||||:|| |||||
Db       415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy       480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
             ||| ||||||||||| |||||| ||||||||||||||  |:||||||||||:|||| |||
Db       475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy       540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
             :||||:| :||||| | : :|: || |    || :: : |   :| | | |||||| :|||
Db       535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy       596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
             ||||||||:| |||||:|||:    :::| |||||||||:|| ||||||:||  ||    |:
Db       595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy       656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
             :|| :|||||::||:||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db       652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy       716 VKMLA 720
             :||||
Db       712 LKMLA 716
```

RESULT 2
US-10-476-198-4
; Sequence 4, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,198
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: C. violaceum
US-10-476-198-4

```
  Query Match            72.4%;  Score 2614;  DB 6;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 3.2e-194;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy         1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
             |||||||:: : |||        |  | |     |      |          :|::||:::|:
Db         1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy        61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
             |||   :|| |||| : ||| |||||| |||:: ||||:::|||:||:||:|| ||||||
Db        55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy       121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
             ||||||||||||:||||||:||||||| |:|  || :||||||:|||||||||| ||||
Db       115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy       181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
             |||||||||||||||||||||| ||||||||:|||||||||||||||| ||:|||| |||
Db       175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy       241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
             |:||| |||::||||::|| || |||||||||||||||||||||||||:|:|:||||||||
Db       235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy       301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
             :|||:||||||  ||| ||:||| |||| |||||||:||| |||||| |||||||||||
Db       295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy       361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
             ||||||||||||||||:||||| |::||::|::|      :|||||:||||| : ||||||
Db       355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy       420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
             |:||: :|||||| || ||||||| ||:||| ||||||||||||||||:|| |||||
Db       415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy       480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
             ||| ||||||||||| |||||| ||||||||||||||  |:||||||||||:|||| |||
Db       475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy       540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
             :||||:| :||||| | : :|: || |    || :: : |   :| | | |||||| :|||
Db       535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy       596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
             ||||||||:| |||||:|||:    :::| |||||||||:|| ||||||:||  ||    |:
Db       595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy       656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
             :|| :|||||::||:||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db       652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy       716 VKMLA 720
             :||||
Db       712 LKMLA 716
```

RESULT 3
US-10-476-193-19
; Sequence 19, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-193-19

```
  Query Match            72.4%;  Score 2613;  DB 6;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 3.8e-194;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy         1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
             |||||||:: : |||        |  | |     |      |          :|::||:::|:
Db         1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy        61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
```

```
Db          55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy         121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db         115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy         181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db         175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy         241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db         235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db         295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db         355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy         420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db         415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy         480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db         475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy         540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db         535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy         596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db         595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy         656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db         652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy         716 VKMLA 720
Db         712 LKMLA 716
```

RESULT 4
US-10-476-198-19
; Sequence 19, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,198
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-198-19

```
  Query Match             72.4%;  Score 2613;  DB 6;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 3.8e-194;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy           1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db           1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy          61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db          55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy         121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db         115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy         181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db         175 SLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy         241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db         235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy         301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db         295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy         361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db         355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy         420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db         415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy         480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db         475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy         540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db         535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy         596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db         595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy         656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db         652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy         716 VKMLA 720
Db         712 LKMLA 716
```

RESULT 5
US-10-476-193-24
; Sequence 24, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid

```
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/476,193
;  CURRENT FILING DATE:  2003-10-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: artificial sequence
;   FEATURE:
;   OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-476-193-24

  Query Match            72.4%;  Score 2612;  DB 6;  Length 716;
  Best Local Similarity  72.0%;  Pred. No. 4.6e-194;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
Db        712 LKMLA 716

RESULT 6
US-10-476-198-24
; Sequence 24, Application US/10476198
```

```
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/476,198
;  CURRENT FILING DATE:  2003-10-24
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: artificial sequence
;   FEATURE:
;   OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-476-198-24

  Query Match            72.4%;  Score 2612;  DB 6;  Length 716;
  Best Local Similarity  72.0%;  Pred. No. 4.6e-194;
  Matches  522;  Conservative   85;  Mismatches  104;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
Db        535 IDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
```

```
Db        712 LKMLA 716

RESULT 7
US-10-476-193-22
; Sequence 22, Application US/10476193
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/476,193
;  CURRENT FILING DATE:  2003-10-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: x=Gly, Ala, Ser,Thr, Pro
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: x=Thr, Ala, Ser, Pro, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-193-22

  Query Match             72.2%;  Score 2607;  DB 6;  Length 716;
  Best Local Similarity   71.9%;  Pred. No. 1.1e-193;
  Matches  521;  Conservative   85;  Mismatches  105;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |    |      |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| |||| :  ||| |||||| |||::  |||:::|||:|||::||  |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|   || :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :||:||||||| ||| ||:|||:||| ||||  |:|||:||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||||:|||||| |::||:: |::|: :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||||  ||||||||| ||:|||  |||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| ||| |||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||  | :||||| | : :|: || |   || :: :  |   :| | | ||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716

RESULT 8
US-10-476-198-22
; Sequence 22, Application US/10476198
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/476,198
;  CURRENT FILING DATE:  2003-10-24
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: x=Gly, Ala, Ser,Thr, Pro
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: x=Thr, Ala, Ser, Pro, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-22

  Query Match             72.2%;  Score 2607;  DB 6;  Length 716;
  Best Local Similarity   71.9%;  Pred. No. 1.1e-193;
  Matches  521;  Conservative   85;  Mismatches  105;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |    |      |        :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||   :|| |||| :  ||| |||||| |||::  |||:::|||:|||::||  |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:||||||| |:|   || :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||||:|||:||||||||
```

```
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419

Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479

Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539

Db        475 AAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595

Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655

Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715

Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720

Db        712 LKMLA 716
```

RESULT 9
US-10-476-193-20
; Sequence 20, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (181)..(181)
; OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (235)..(235)
; OTHER INFORMATION: X=Ala, Gly, Ser, Thr, Pro
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-193-20

Query Match 72.1%; Score 2603; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 2.3e-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

```
Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120

Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180

Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240

Db        175 SLTRGHXAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300

Db        235 XVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360

Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419

Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479

Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539

Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595

Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655

Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715

Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720

Db        712 LKMLA 716
```

RESULT 10
US-10-476-193-21
; Sequence 21, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis

```
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (149)..(149)
;   OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (202)..(202)
;   OTHER INFORMATION: X=Val, Met, Leu, Cys
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-193-21

  Query Match             72.1%;  Score 2603;  DB 6;  Length 716;
  Best Local Similarity   71.7%;  Pred. No. 2.3e-193;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||       | | |     |       |       :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: |||:::||||:|||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||||  |:|     | :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDXFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||| ||| |||||:|||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPXVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::|||||::||| || ||||||||||||||||||||||||:|||:||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:||||| |::||:: |::|    :||||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || |||||||||| ||:|||||||||||||||||||||:|||||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| ||||||| |||||||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :|||||| | : :|: || |    || :: : |  :| | | |||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| |||||||||:|| ||||||:|| ||   |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||||::|:||||||||:||||||||: |:|||||||||:|||||:||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

```
RESULT 11
US-10-476-198-20
; Sequence 20, Application US/10476198
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/476,198
;  CURRENT FILING DATE:  2003-10-24
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (181)..(181)
;   OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (235)..(235)
;   OTHER INFORMATION: X=Ala, Gly, Ser, Thr, Pro
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-20

  Query Match             72.1%;  Score 2603;  DB 6;  Length 716;
  Best Local Similarity   71.7%;  Pred. No. 2.3e-193;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||       | | |     |       |       :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: |||:::||||:|||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||||  |:|  || :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||| |||||||||||||||| ||||||||:|||||||||||||| ||:|||| |||
Db        175 SLTRGHXAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              :||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||
Db        235 XVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||:||||| |::||:: |::|    :||||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || |||||||||| ||:|||||||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474
```

```
Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| |||||| ||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :||||| | : :|: || |    || :: : |   :| | | |||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: ||||||:||||||  ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 12
US-10-476-198-21
; Sequence 21, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,198
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X=Val, Met, Leu, Cys
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-21

```
  Query Match             72.1%;  Score 2603;  DB 6;  Length 716;
  Best Local Similarity   71.7%;  Pred. No. 2.3e-193;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        | | |      |      |          :|::|||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| ||||| |||:: ||||:::||||:||:||:| |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||||:||||||:||||||| |:|   | :||||||:||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDXFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| ||| |||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPXVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||| ||| ||:||| |||| |:||| ||| || ||  |||  |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:|||||| |::||:: |::|    :||||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :|||||| || ||||||| ||:||| |||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| |||||| ||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :||||| | : :|: || |    || :: : |   :| | | |||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: ||||||:||||||  ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 13
US-10-476-193-18
; Sequence 18, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (126)..(126)
; OTHER INFORMATION: X=Gly, Ala, Ser, Thr
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (138)..(138)
; OTHER INFORMATION: X=Leu, Met, Ile, Val, Cys
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-193-18

```
  Query Match             72.0%;  Score 2599;  DB 6;  Length 716;
  Best Local Similarity   71.7%;  Pred. No. 4.7e-193;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||:: : |||        | | |    |       |       :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| ||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||| |:||||||:|| |||| |:|  || :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEXAISLQKALLEHXLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||||| ||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || |||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||||:||||| |::||:: |::|   :||||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| ||||||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :|||||  | : :|: || |    || :: : |  :| | | ||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              |||||||:| |||||:|||:    :::| ||||||||:||| ||||||:|| ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:|||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

```
RESULT 14
US-10-476-198-18
; Sequence 18, Application US/10476198
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/476,198
;  CURRENT FILING DATE:  2003-10-24
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (126)..(126)
;   OTHER INFORMATION: X=Gly, Ala, Ser, Thr
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (138)..(138)
;   OTHER INFORMATION: X=Leu, Met, Ile, Val, Cys
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-198-18
```


```
  Query Match             72.0%;  Score 2599;  DB 6;  Length 716;
  Best Local Similarity   71.7%;  Pred. No. 4.7e-193;
  Matches  520;  Conservative   85;  Mismatches  106;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              ||||||:: : |||        | | |    |       |       :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| ||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||| |:||||||:|| |||| |:|  || :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEXAISLQKALLEHXLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||||| ||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || |||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||| ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||||:||||| |::||:: |::|   :||||||:||||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| ||||||||||||||| |:||||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :|||| | :|||||  | : :|: || |    || :: : |  :| | | ||||| :|||
Db        535 IDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              |||||||:| |||||:|||:    :::| ||||||||:||| ||||||:|| ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:|||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

```
RESULT 15
US-10-476-193-17
; Sequence 17, Application US/10476193
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT: Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION: hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/476,193
;  CURRENT FILING DATE: 2003-10-28
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (126)..(126)
;   OTHER INFORMATION: X= Gly, Ser, Ala, Thr
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (138)..(138)
;   OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (149)..(149)
;   OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (181)..(181)
;   OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: X=Asp, Asn, Glu, Gln
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (202)..(202)
;   OTHER INFORMATION: X=Val, Met, Leu, Cys
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (235)..(235)
;   OTHER INFORMATION: X=Ala, Gly, Ser, Pro
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: X=Gly, Ala, Ser, Thr, Pro
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-193-17

  Query Match             71.4%;  Score 2579;  DB 6;  Length 716;
  Best Local Similarity   71.0%;  Pred. No. 1.7e-191;
  Matches  515;  Conservative   86;  Mismatches  110;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||       |  |   |     |         :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| ||||  : ||| |||||| |||::  ||||:::||||:||::||  |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||| |:||||||:|| |||| |:|    | :|||||:|||||||||| ||||
Db        115 GFGGSADTRTEXAISLQKALLEHXLCGVLPSSFDXFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||| |||||||||||||||| ||| |||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHXAVRLVVLEALTNFLNHXITPXVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
               :||| |||::||||::|| || |||||||||||||||||||||||:|||:|||||||
Db        235 XVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||   ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              ||||||||||||||||:|||||  |::||:: |::|    :||||||:|||| :  |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || ||||||| ||:||| ||||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| |||||| ||| |||||||||| |:||||||||||:||||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :||||| | : :|: || |     || :: : |  :| | | |||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:     :::| ||||||||:|| ||||||:|| || |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||::||:||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716


Search completed: September  9, 2004, 10:21:12
Job time : 46.8174 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:          September  9, 2004, 09:58:38 ; Search time 28.8797 Seconds
                                               (without alignments)
                                               2398.154 Million cell updates/sec

Title:           US-09-939-408A-13
Perfect score:   3610
Sequence:        1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        PIR_78:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2614	72.4	716	2	A56628	phenylalanine ammo
2	2573	71.3	693	2	A29607	phenylalanine ammo
3	2499.5	69.2	713	2	S01999	phenylalanine ammo
4	1041.5	28.9	725	2	G84787	phenylalanine ammo
5	1035.5	28.7	725	2	S52990	phenylalanine ammo
6	1017.5	28.2	718	2	S48726	phenylalanine ammo
7	1016.5	28.2	715	2	S66343	phenylalanine ammo
8	1016	28.1	715	2	JQ2265	phenylalanine ammo
9	1015.5	28.1	717	2	T46172	phenylalanine ammo
10	1011	28.0	713	2	S22991	phenylalanine ammo
11	1009	28.0	725	2	S17444	phenylalanine ammo
12	1006	27.9	716	2	S04463	phenylalanine ammo
13	1005.5	27.9	723	2	S25303	phenylalanine ammo
14	1005	27.8	712	2	T03663	phenylalanine ammo
15	1005	27.8	716	2	S48725	phenylalanine ammo
16	1004	27.8	701	2	S06475	phenylalanine ammo
17	1003.5	27.8	712	2	S04127	phenylalanine ammo
18	998	27.6	705	2	JC5873	phenylalanine ammo
19	997	27.6	720	2	S21174	phenylalanine ammo
20	995	27.6	710	2	JC5872	phenylalanine ammo
21	994.5	27.5	721	2	A44133	phenylalanine ammo
22	993.5	27.5	708	2	T10909	phenylalanine ammo
23	993	27.5	712	2	T01858	phenylalanine ammo
24	987.5	27.4	667	2	T12749	phenylalanine ammo
25	983	27.2	708	2	T14295	phenylalanine ammo
26	982.5	27.2	710	2	S66313	phenylalanine ammo
27	973.5	27.0	700	2	T06545	phenylalanine ammo
28	973.5	27.0	754	2	T09777	phenylalanine ammo
29	966	26.8	590	2	S70916	phenylalanine ammo
30	964.5	26.7	695	2	S52992	phenylalanine ammo
31	955	26.5	549	2	T05970	phenylalanine ammo
32	954.5	26.4	707	2	S29029	phenylalanine ammo
33	953.5	26.4	710	2	S60042	phenylalanine ammo
34	946.5	26.2	710	2	S04128	phenylalanine ammo
35	946	26.2	619	2	S52632	phenylalanine ammo
36	934	25.9	716	2	S28185	phenylalanine ammo
37	875.5	24.3	582	2	S60043	phenylalanine ammo
38	703.5	19.5	506	2	S04129	phenylalanine ammo
39	697.5	19.3	505	2	A24727	phenylalanine ammo
40	659.5	18.3	497	2	T05968	probable phenylala
41	655.5	18.2	475	2	T05966	phenylalanine ammo
42	547.5	15.2	524	2	F75610	histidine ammonia-
43	543.5	15.1	416	2	JQ1070	phenylalanine ammo
44	543	15.0	514	2	AG3040	histidine ammonia-
45	543	15.0	514	2	E98245	hutH gene homolog

ALIGNMENTS

RESULT 1
A56628
phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodosporidium toruloides) (strain CB
C;Species: Rhodosporidium toruloides
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C;Accession: A56628
R;Rasmussen, O.F.; Oerum, H.
DNA Seq. 1, 207-211, 1991
A;Title: Analysis of the gene for phenylalanine ammonia-lyase from Rhodosporidium torulo
A;Reference number: A56628; MUID:92127061; PMID:1773059
A;Accession: A56628
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-716 <RAS>
A;Cross-references: GB:X51513; NID:g3293; PIDN:CAA35886.1; PID:g3294
A;Note: sequence extracted from NCBI backbone (NCBIP:78381)
C;Genetics:
A;Gene: pal
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;212/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             72.4%;  Score 2614;  DB 2;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 1.3e-165;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||       |  | |    |      |        :|::||:::|: 
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||::||  |||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              |||||||||||||:||||||:||||||| |:|  || :||||||:|||||||||| ||||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              ||||||||||||||||||||||| |||||||||:|||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:|||||||  ||| ||:||| |||| |:||| ||| || || ||| ||||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
```

```
              |||||||||||||||:||||| |::||:: |::|   :||||||:||||| : ||||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || |||||||| ||:||| ||||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| |||||||||||| |||||| |||||||||||||| |:||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :|||||| | : :|: || |   || :: : |   :| | | ||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:|| ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :||||||::||:||||||||:|||||||:| |: |||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716
```

RESULT 2
A29607
phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodosporidium toruloides) (strain IF
C;Species: Rhodosporidium toruloides
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 16-Jul-1999
C;Accession: A29607
R;Anson, J.G.; Gilbert, H.J.; Oram, J.D.; Minton, N.P.
Gene 58, 189-199, 1987
A;Title: Complete nucleotide sequence of the Rhodosporidium toruloides gene coding for p
A;Reference number: A29607; MUID:88112870; PMID:2828184
A;Accession: A29607
A;Molecule type: DNA
A;Residues: 1-693 <ANS>
A;Cross-references: GB:M18261; NID:g169745; PIDN:AAA33883.1; PID:g169746
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;188-190/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;189/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             71.3%;  Score 2573;  DB 2;  Length 693;
  Best Local Similarity   74.9%;  Pred. No. 6.5e-163;
  Matches  507;  Conservative   83;  Mismatches   79;  Indels    8;  Gaps    4;

Qy         49 SQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLK 108
              :|::||:::|: |||  :|| |||| : ||| |||||| |||:: ||||:::||||:||:
Db         20 TQVDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLR 79

Qy        109 AQLQNSVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPL 168
              :||  |||||||||||||||||||:||||||:|||||||| |:|  || :||||||:|||
Db         80 SQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPL 139

Qy        169 EVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIA 228
              |||||| |||||||||||||||||||||||||||| |||||||||:||||||||||||||
Db        140 EVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIA 199

Qy        229 GAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLS 288
              ||:|||| ||||:||| |||::||||||||:|||||||||||||||||||||||||||||:
Db        200 AAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLA 259

Qy        289 LHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFA 348
              |||:||||||||:|||:|||||||  ||| |:|||  |||| |:||| ||| || || ||
Db        260 LHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFA 319

Qy        349 VEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDV 407
              | ||||||||||||||||||||||||||:||||| |::||:: |::|   :||||||:||

Db        320 VHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDV 379

Qy        408 ENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPS 467
              ||| : ||||||:||: :|||||| || ||||||||| ||:||| ||||||||||||||
Db        380 ENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPS 439

Qy        468 LNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSL 527
              |:|| |||||  ||| |||||||||||| |||||| |||||||||||||| |:||||||
Db        440 LSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSL 499

Qy        528 LLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALN 583
              |||:|||| |||:||||:|: |||||| | : :|: ||      || :: : |  :| | |
Db        500 LLATHLYCVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLA 559

Qy        584 KRLEQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTRE 643
              |||||| :||| ||||||||:| |||||:|||:    :::| ||||||||:|| ||||||:
Db        560 KRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQ 616

Qy        644 VRNRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIY 703
              || ||    |: :|| :||||||::||:||||||||:|||||||:| |: |||||||:||
Db        617 VRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIY 676

Qy        704 EAIKDGRINHVLVKMLA 720
              |||| ||||:||:||||
Db        677 EAIKSGRINNVLLKMLA 693
```

RESULT 3
S01999
phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodotorula rubra)
C;Species: Rhodotorula rubra
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
C;Accession: S01999; S01990
R;Filpula, D.; Vaslet, C.A.; Levy, A.; Sykes, A.; Strausberg, R.L.
Nucleic Acids Res. 16, 11381, 1988
A;Title: Nucleotide sequence of gene for phenylalanine ammonia-lyase from Rhodotorula r
A;Reference number: S01999; MUID:89083582; PMID:3205749
A;Accession: S01999
A;Molecule type: DNA
A;Residues: 1-713 <FIL>
A;Cross-references: EMBL:X13094; NID:g3284; PIDN:CAA31486.1; PID:g295942
R;Vaslet, C.A.; Strausberg, R.L.; Sykes, A.; Levy, A.; Filpula, D.
Nucleic Acids Res. 16, 11382, 1988
A;Title: cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes reveal ger
A;Reference number: S01990; MUID:89083583; PMID:3205750
A;Accession: S01990
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-674 <VAS>
A;Cross-references: EMBL:X13095
C;Genetics:
A;Introns: 110/2; 139/2; 372/3; 428/2; 482/3
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;217-219/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;218/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             69.2%;  Score 2499.5;  DB 2;  Length 713;
  Best Local Similarity   69.6%;  Pred. No. 5.2e-158;
  Matches  511;  Conservative   83;  Mismatches  105;  Indels   35;  Gaps    9;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPT-------KSAAGPTSALRRTPGLDGHAAHQSQLEI 53
              ||||:||:||::||  :|| |||        |  ||  | |  |           :||:|
Db          1 MAPSVDSIATSVANSLSNGLHAAAAANGGDVHKKTAGAGSLLPTT--------ETTQLDI 52

Qy         54 VQELLSDP-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQ 112
              |: :|:|      | ::| ||:||: ||||||:|| |:| :    || ::| ||:||: ||
Db         53 VERILADAGATDQIKLDGYTLTLGDVVGAARRGRSVKVADSPHIREKIDASVEFLRTQLD 112

Qy        113 NSVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVR 172
              ||||||||||||||||||||:||||||:||||||||| |||  :::|||||||:||||||
Db        113 NSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSMDGFALGRGLENSLPLEVVR 172
```

```
Qy        173 GAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAIT 232
Db        173 GAMTIRVNSLTRGHSAVRIVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAASIT 232

Qy        233 GHPDVKVHVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDS 292
Db        233 GHPDSKVHV--DG--KIMSAQEAIALKGLQPVVLGPKEGLGLVNGTAVSASMATLALTDA 288

Qy        293 HMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHE 352
Db        289 HVLSLLAQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSKYAVHHE 348

Qy        353 EEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQ 411
Db        349 TEVKVKDDEGILRQDRYPLRCSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDLENKM 408

Qy        412 TAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYH 471
Db        409 THHGGAFMASSVGNTMEKTRLAVALMGKVSFTQLTEMLNAGMNRALPSCLAAEDPSLSYH 468

Qy        472 GKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLAS 531
Db        469 CKGLDIAAAAYTSELGHLANPVSTHVQPAEMGNQAINSLALISARRTAEANDVLSLLLAT 528

Qy        532 HLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALAL-----EVKKALNKRL 586
Db        529 HLYCVLQAVDLRAMEFEHTKAFEPMVTELLKQHFG------ALATAEVEDKVRKSIYKRL 582

Qy        587 EQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRN 646
Db        583 QQNNSYDLEQRWHDTFSVATGAVVEALAGQ---EVSLASLNAWKVACAEKAIALTRSVRD 639

Qy        647 RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAI 706
Db        640 SFWAAPSSSSPALKYLSPRTRVLYSFVREEVGVKARRGDVYLGKQEVTIGTNVSRIYEAI 699

Qy        707 KDGRINHVLVKMLA 720
Db        700 KSGCIAPVLVKMMA 713
```

RESULT 4
G84787
phenylalanine ammonia lyase (PAL1) [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: G84787
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-725 <STO>
A;Cross-references: GB:AE002093; NID:g6598547; PIDN:AAD18156.2; GSPDB:GN00139
C;Genetics:
A;Gene: At2g37040
A;Map position: 2
C;Superfamily: histidine ammonia-lyase

```
  Query Match             28.9%;  Score 1041.5;  DB 2;  Length 725;
  Best Local Similarity   36.5%;  Pred. No. 3.7e-61;
  Matches  272;  Conservative  138;  Mismatches  258;  Indels   77;  Gaps   21;

Qy         18 NGSHAAPTKSAAGPTSAL--------------RRTPGLDGHAAHQ---SQLEIVQELLSD 60
Db          4 NGAH----KSNGGGVDAMLCGGDIKTKNMVINAEDPLNWGAAAEQMKGSHLDEVKRMVAE 59

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGV 118
Db         60 FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARAGVNASSDWVMESMNKGTDSYGV 119

Qy        119 TTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIR 178
Db        120 TTGFGATSHRRTKNGVALQKELIRFLNAGI-------FGSTKETSHTLPHSATRAAMLVR 172

Qy        179 VNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVK 238
Db        173 INTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTITASGDLVPLSYIAGLLTGRPNSK 232

Qy        239 VHVLHEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHML 295
Db        233 A----TGPNGEALTAEEAFKLAGISSGFFDLQPKEGLALVNGTAVGSGMASMVLFETNVL 288

Qy        296 SLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEV 355
Db        289 SVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAIMEHILDGSSY-MKLAQKL 346

Qy        356 KVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHG 415
Db        347 HEMDPLQKPKQDRYALRTSPQWLGPQIEVIRYATKSIEREINSVNDNPLIDVSRNKAIHG 406

Qy        416 GNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKG 474
Db        407 GNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKG 466

Qy        475 LDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLY 534
Db        467 AEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLV 526

Qy        535 CTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRL------EQ 588
Db        527 AICQAVDLRHLEENLRQTVKNTVSQVAKKVLTTG--VNG-ELHPSRFCEKDLLKVVDREQ 583

Qy        589 TTTYDLEPRWHDAFSYATGTVVE----------LLSSSPSANVTLTAVNAWKVASAEKAI 638
Db        584 VYTYADDP------CSATYPLIQKLRQVIVDHALINGESEKNAVTSIFH--KIGAFEEEL 635

Qy        639 S--LTREVR--NRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQET 694
Db        636 KAVLPKEVEAARAAYDNGTSAIP-NRIKECRSYPLYRFVREELGT-----ELLTGEKVTS 689

Qy        695 IGSNVSRIYEAIKDGRINHVLVKML 719
Db        690 PGEEFDKVFTAICEGKIIDPMMECL 714
```

RESULT 5
S52990
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C;Accession: S52990; PQ0140
R;Wanner, L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
Plant Mol. Biol. 27, 327-338, 1995
A;Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
A;Reference number: S52990; MUID:95195160; PMID:7888622
A;Accession: S52990
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-725 <WAN>
A;Cross-references: EMBL:L33677; NID:g497418; PIDN:AAC18870.1; PID:g497419
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
R;Ohl, S.; Hedrick, S.A.; Chory, J.; Lamb, C.J.
Plant Cell 2, 837-848, 1990
A;Title: Functional properties of a phenylalanine ammonia-lyase promoter from Arabidopsi
A;Reference number: PQ0140; MUID:93005677; PMID:2152131
A;Accession: PQ0140
A;Molecule type: DNA
A;Residues: 1-240 <OHL>
A;Cross-references: GB:X62747; NID:g16408; PIDN:CAA44609.1; PID:g16409
A;Experimental source: Ecotype Columbia
C;Comment: Phenylalanine ammonia-lyase catalyzes the deamination of L-phenylalanine to t

```
C;Genetics:
A;Gene: PAL1
A;Introns: 143/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;212/Modified site: dehydroalanine (Ser) #status predicted

  Query Match             28.7%;  Score 1035.5;  DB 2;  Length 725;
  Best Local Similarity   36.4%;  Pred. No. 9.2e-61;
  Matches  271;  Conservative  138;  Mismatches  259;  Indels   77;  Gaps   21;

Qy         18 NGSHAAPTKSAAGPTSAL-------------RRTPGLDGHAAHQ---SQLEIVQELLSD 60
              ||:|    ||  |    |:                |    | ||  |    | |: |: ::::
Db          4 NGAH----KSNGGGVDAMLCGGDIKTKNMVINAEDPLNWGAAAEQMKGSHLDEVKRMVAE 59

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGV 118
                    || |  | :||:  |    :   |  |:|:  :  || |: | |::    :     |||
Db         60 FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARAGVNASSDWVMESMNKGTDSYGV 119

Qy        119 TTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIR 178
              ||||| ::   ||::  |:|||  ||         |:          |      :|||     |  ||::|
Db        120 TTGFGATSHRRTKNGVALQKELIRFLNAGI-------FGSTKETSHTLPHSATRAAMLVR 172

Qy        179 VNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVK 238
              :|:|  :|  |  :|   :|||:|:|||: |||  :||||:|:||||| ||||||  :||  |:  |
Db        173 INTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTITASGDLVPLSYIAGLLTGRPNSK 232

Qy        239 VHVLHEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHML 295
                        |    :  :  |  ||   |  |: :       |  |||||  ||||||  :  ||::  |  ::::|
Db        233 A----TGPNGEALTAEEAFKLAGISSGFFDLQPKEGLALVNGTAVGSGMASMVLFETNVL 288

Qy        296 SLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEV 355
              |:|::  |:|:   |  |  |:      |     :       :  |||:|  |     :  |  |||:  ::   :::
Db        289 SVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAVMEHILDGSSY-MKLAQKL 346

Qy        356 KVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHG 415
                        |       :|||  ||||||:|||  :|  :  :|     ::    |  |:    ||||:||    :   ||
Db        347 HEMDPLQKPKQDRYALRTSPQWLGPQIEVIRYATKSIEREINSVNDNPLIDVSRNKAIHG 406

Qy        416 GNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKG 474
              ||||  :  :  :||:  ||||:   ||||  |  |  :||:|     |  ||||  |  |:  :|||:|   ||
Db        407 GNFQGTPIGVSMDNTRLAIRAIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKG 466

Qy        475 LDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLY 534
                :|  :|:|  |||  :||||||:  ||  ||   ||  ||||  |||:|:|:|:||  |:|  |:   :   |
Db        467 AEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLV 526

Qy        535 CTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRL------EQ 588
              ||||||  :|  ::         :   :  :  ||  ||    |       :         |        ||
Db        527 AICQAVDLRHLEENLRQTVKNTVSQVAKKVLTTG--VNG-ELHPSRFCEKDLLKVVDREQ 583

Qy        589 TTTYDLEPRWHDAFSYATGTVVE----------LLSSSPSANVTLTAVNAWKVASAEKAI 638
                 ||  :|        ||   :::                |::      |     :    :   |:  :  |:  :
Db        584 VYTYADDP------CSATYPLIQKLRQVIVDHALVNGESEKNAVTSIFH--KIGAFEEEL 635

Qy        639 S--LTREVR--NRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQET 694
                  |  :||        :       :|  |   |:     |:    ||  ||||||||         ::   |  :   :
Db        636 KAVLPKEVEAARAAYDNGTSAIP-NRIKECRSYPLYRFVREELGT-----ELLTGEKVTS 689

Qy        695 IGSNVSRIYEAIKDGRINHVLVKML 719
                 |       :::  ||  :|:|       :::  |
Db        690 PGEEFDKVFTAICEGKIIDPMMECL 714

RESULT 6
S48726
phenylalanine ammonia-lyase (EC 4.3.1.5) 3 - parsley
C;Species: Petroselinum crispum (parsley)
C;Date: 18-Oct-1995 #sequence_revision 12-Apr-1996 #text_change 16-Jul-1999
C;Accession: S48726; S56035
R;Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
Eur. J. Biochem. 225, 491-499, 1994
```

```
A;Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase i
A;Reference number: S48725; MUID:95010141; PMID:7925471
A;Accession: S48726
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-718 <APP1>
A;Cross-references: EMBL:X81159
R;Appert, C.
submitted to the EMBL Data Library, August 1994
A;Reference number: S56035
A;Accession: S56035
A;Molecule type: mRNA
A;Residues: 1-58,'E',60-718 <APP2>
A;Cross-references: EMBL:X81159; NID:g535007; PIDN:CAA57057.1; PID:g535008
C;Genetics:
A;Gene: PAL3
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;204-206/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;205/Modified site: dehydroalanine (Ser) #status predicted

  Query Match             28.2%;  Score 1017.5;  DB 2;  Length 718;
  Best Local Similarity   36.5%;  Pred. No. 1.4e-59;
  Matches  269;  Conservative  128;  Mismatches  275;  Indels   65;  Gaps   19;

Qy         14 NGFTNGSHAAPTKSAAGPTSALRRTP----GLDGHAAHQSQLEIVQELLSDPTDDVVELS 69
              || |||  ||      :  |      :::          |:     |       |  |:  |: ::::       ||:|
Db          5 NGTTNG-HA----NGNGLDLCMKKEDPLNWGVAAEALTGSHLDEVKRMVAEYRKPVVKLG 59

Qy         70 GYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSA 126
               | :||:    :||    |:|:  :|  ||  |   |  | |::     :          ||||||||  ::
Db         60 GETLTISQVAAISARDDSGVKVELSEEARAGVKASSDWVMDSMNKGTDSYGVTTGFGATS 119

Qy        127 DTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE---NTLPLEVVRGAMVIRVNSLT 183
                ||:     :||| ||    |:                 |  |  |      ||||       | ||::|:|:|
Db        120 HRRTKQGGALQKELIRFLNAGI---------FGSGAEAGNNTLPHSATRAAMLVRINTLL 170

Qy        184 RGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLH 243
              :|:|  :|    :|||:|  ||||  |||  :||||:|:||||| ||||||  :||  |:   |
Db        171 QGYSGIRFEILEAITKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA---V 227

Qy        244 EGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQA 301
                  |     :     ||   |   | |:|          |  |||||  ||||||  :  ||::  |  ::::|::|::
Db        228 GPTGVTLSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEV 287

Qy        302 LTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDE 361
              ::|:   |  |:   |     |         :   ||||:|   |:  :  ||||:   |:   ::
Db        288 MSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEMDPL 345

Qy        362 GILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQAS 421
                 :||||  ||||||:|||  :|  :    :     :   |  |:    ||||:||     :   ||||||  |
Db        346 QKPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGS 405

Qy        422 AVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIA 480
                :  :||:  ||||:|   ||||  |  |  :||:|     |  ||||  |:     :|||:|   ||  :|  :|
Db        406 PIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMA 465

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540
              :|  ||  |||||||   ||  ||   ||  ||||  |||:|:|:|:||  :|:|  |:   :       ||:
Db        466 SYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAI 525

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ----------- 588
              ||| :| :  :            :   :  ::  |      |  ||      |         |  |   :
Db        526 DLRHLEENLKSTVKNTVSQVAKRVLTMG--VNG-ELHPSRFCEKDLLRVVDREYIFAYID 582

Qy        589 ---TTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS--LTRE 643
                  :  ||  |   :   :                |:||     ::          |:          |:|:  |    :      |  :|
Db        583 DPCSATYPLMQKLRE-------TLVEHALNNGDKERNLSTSIFQKIAAFEDELKALLPKE 635

Qy        644 VRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRI 702
                |              :  |   :         |:    ||  ||||||||          :        :      :        |           ::
Db        636 VETARAALESGNPAIPNRIKECRSYPLYKFVREELGT-----EYLTGEKVRSPGEEFEKV 690
```

```
Qy        703 YEAIKDGRINHVLVKML 719
              : |:  | |   |:: |
Db        691 FTAMSKGEIIDPLLECL 707


RESULT 7
S66343
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999
C;Accession: S66343
R;Fukasawa-Akada, T.; Kung, S.; Watson, J.C.
Plant Mol. Biol. 30, 711-722, 1996
A;Title: Phenylalanine ammonia-lyase gene structure, expression, and evolution in Nicoti
A;Reference number: S66343; MUID:96194455; PMID:8624404
A;Accession: S66343
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-715 <FUK>
A;Cross-references: EMBL:M84466; NID:g170349; PIDN:AAA34122.1; PID:g170350
C;Genetics:
A;Gene: PAL1
A;Introns: 133/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;201-203/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;202/Modified site: dehydroalanine (Ser) #status predicted

  Query Match            28.2%;  Score 1016.5;  DB 2;  Length 715;
  Best Local Similarity  36.0%;  Pred. No. 1.6e-59;
  Matches  270;  Conservative  130;  Mismatches  259;  Indels   91;  Gaps   19;

Qy         13 ANGFTN-GSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVELSGY 71
Db          3 SNGHVNGGENFELCKKSADPLNWEMAAESLRG-----SHLDEVKKMVSEFRKPMVKLGGE 57

Qy         72 SLTVRDVVGAA---RKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSA 126
Db         58 SLTVAQVAAIAVRDKSANGVKVELSEEARAGVKASSDWVMDSMNKGTDSYGVTTGFGATS 117

Qy        127 DTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGH 186
Db        118 HRRTKNGGALQKELIRFLNAGV-------FGNGTETSHTLPHSATRAAMLVRINTLLQGY 170

Qy        187 SAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGT 246
Db        171 SGIRFEILEAITKLINSNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPNG- 228

Qy        247 EKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTA 304
Db        229 -ETLNAEEAFRVAGVNGGFFELQPKEGLALVNGTAVGSGMASMVLFDSNILAVMSEVLSA 287

Qy        305 LTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGIL 364
Db        288 IFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSY-VKAAQKLHEMDPLQKP 345

Qy        365 RQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVS 424
Db        346 KQDRYALRTSPQWLGPQIEVIRAATKMIEREINSVNDNPLIDVSRNKALHGGNFQGTPIG 405

Qy        425 ISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGLDIHIAAYA 483
Db        406 VSMDNARLALASIGKLMFAQFSELVNDYYNNGLPSNLTASRNPSLDYGFKGAEIAMASYC 465

Qy        484 SELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLR 543
Db        466 SELQFLANPVTNHVQSAEQHNQDVNSLGLISARKTAEAVDILKLMSSTYLVALCQAIDLR 525

Qy        544 AMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDA-- 601
Db        526 HLEENLK---------------------NA----VKNTVSQVAKRTLTMGANGELHPARF 560

Qy        602 --------------FSYATG--------------TVVELLSSSPSANVTLTAVNAWKVAS 633
Db        561 CEKELLRIVDREYLFAYADDPCSCNYPLMQKLRQVLVDHAMNNGESEKNVNSSIFQKIGA 620

Qy        634 AE---KAISLTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVG 689
Db        621 FEDELKAV-LPKEVESARAALESGNPAIPNRITECRSYPLYRFVRKELGT-----ELLTG 674

Qy        690 VQQETIGSNVSRIYEAIKDGRINHVLVKML 719
Db        675 EKVRSPGEECDKVFTAMCNGQIIDPMLECL 704


RESULT 8
JQ2265
phenylalanine ammonia-lyase (EC 4.3.1.5) - western balsam poplar x cottonwood
C;Species: Populus trichocarpa x Populus deltoides (western balsam poplar x cottonwood)
C;Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 28-May-1999
C;Accession: JQ2265
R;Subramaniam, R.; Reinold, S.; Molitor, E.K.; Douglas, C.J.
Plant Physiol. 102, 71-83, 1993
A;Title: Structure, inheritance, and expression of hybrid poplar (Populus trichocarpa x
A;Reference number: JQ2265; MUID:94151434; PMID:8108506
A;Accession: JQ2265
A;Molecule type: mRNA
A;Residues: 1-715 <SUB>
A;Cross-references: GB:L11747; NID:g169453; PIDN:AAA33805.1; PID:g169454
A;Note: the authors translated the codon AAG for residue 331 as GLu and AAG for residue
C;Comment: The enzyme plays a role in linking primary metabolism to phenylpropanoid meta
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;201-203/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;202/Modified site: dehydroalanine (Ser) #status predicted

  Query Match            28.1%;  Score 1016;  DB 2;  Length 715;
  Best Local Similarity  35.8%;  Pred. No. 1.8e-59;
  Matches  266;  Conservative  132;  Mismatches  276;  Indels   68;  Gaps   18;

Qy          8 LATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVE 67
Db          1 METVTKNGYQNGSLESLCVNQRDPLSW-----GVAAEAMKGSHLDEVKRMVADYRKPVVK 55

Qy         68 LSGYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGG 124
Db         56 LGGETLTIAQVASIAGHDTGDVKVELSESARPGVKASSDWVMDSMDKGTDSYGVTTGFGA 115

Qy        125 SADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE--NTLPLEVVRGAMVIRVNSL 182
Db        116 TSHRRTKQGGALQKELIRFLNAGI---------FGNGTETCHTLPHSATRAAMLVRINTL 166

Qy        183 TRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVL 242
Db        167 LQGYSGIRFEILEAITRLLNNNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA--- 223

Qy        243 HEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
Db        224 TGPTGEVLDAAEAFKAAGIESGFFELQPKEGLALVNGTAVGSGLASMVLFETNVLAVLSE 283

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVE----HEEEVK 356
Db        284 LLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAYMKAAKKLHETDPL 342

Qy        357 VKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGG 416
Db        343 QKP-----KQDRYALRTSPQWLGPQIEVIRFSTKSIEREINSVNDNPLIDVSRNKAIHGG 397

Qy        417 NFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGL 475
Db        398 NFQGTPIGVSMDNVRLAIASIGKLLFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKGA 457

Qy        476 DIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYC 535
Db        458 EIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSTTFLVA 517
```

```
Qy        536 TLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ------- 588
              ||:||| :| : |         :  : :: | || :     |    :   | | :
Db        518 LCQAIDLRHLEENLKSAVKNTVSQVSKRVLTTGANGE---LHPSRFCEKELLKVVDREYV 574

Qy        589 --------TTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS- 639
                      : || |  :    |       |    | :  :       |:|    |: : |: :
Db        575 FAYVDDPCSATYPLMQKLRQVF------VDHALENGENEKNFSTSV-FQKIEAFEEELKA 627

Qy        640 -LTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGS 697
               | :|| : |    : |  :      |:   || ||||||          :   | : :: |
Db        628 LLPKEVESARAAYDSGNSAIDNKIKECRSYPLYKFVREELGTV-----LLTGEKVQSPGE 682

Qy        698 NVSRIYEAIKDGRINHVLVKML 719
                ::: |:  |:|   ::: |
Db        683 EFDKVFTAMCQGKIIDPMLECL 704
```

RESULT 9
T46172
phenylalanine ammonia-lyase (EC 4.3.1.5) 2 [similarity] - Arabidopsis thaliana
N;Alternate names: protein T4D2.190
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Mar-2000
C;Accession: T46172; S52991
R;Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23025
A;Accession: T46172
A;Molecule type: DNA
A;Residues: 1-717 <NYA>
A;Cross-references: EMBL:AL132958; NID:g6434245; PIDN:CAB64229.1; PID:g6630746
A;Experimental source: cultivar Columbia; BAC clone T4D2
R;Wanner, L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
Plant Mol. Biol. 27, 327-338, 1995
A;Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
A;Reference number: S52990; MUID:95195160; PMID:7888622
A;Accession: S52991
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-205,'H',207-234,'E',236-717 <WAN>
A;Cross-references: EMBL:L33678; NID:g497420; PIDN:AAC18871.1; PID:g497421
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
C;Genetics:
A;Gene: PAL2
A;Map position: 3
A;Introns: 135/2
A;Note: T4D2.190
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;203-205/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;204/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.1%;  Score 1015.5;  DB 2;  Length 717;
  Best Local Similarity   36.0%;  Pred. No. 1.9e-59;
  Matches  264;  Conservative  140;  Mismatches  284;  Indels   45;  Gaps   17;

Qy          5 LDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDD 64
               :| :     |         |  ||: | | :        ||         | |: |:::: :
Db          1 MDQIEAMLCGGGEKTKVAVTTKTLADPLNW-----GLAADQMKGSHLDEVKKMVEEYRRP 55

Qy         65 VVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGF 122
              || | | :||:  |    :  |  |:|:  :  || |   | |::   :      ||||||
Db         56 VVNLGGETLTIGQVAAISTVGGSVKVELAETSRAGVKASSDWVMESMNKGTDSYGVTTGF 115

Qy        123 GGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSL 182
              | ::  ||::  :||   ||       |:       :    :|||    | ||::|||:|
Db        116 GATSHRRTKNGTALQTELIRFLNAGI-------FGNTKETCHTLPQSATRAAMLVRVNTL 168

Qy        183 TRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVL 242
               :|:| :|  :|||:|: ||| |:| :||||:|||||  |||||| :|| |: |
Db        169 LQGYSGIRFEILEAITSLLNHNISPSLPLRGTITASGDLVPLSYIAGLLTGRPNSKA--- 225

Qy        243 HEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
              | : : : |:||    |:     |  ||||| |||||||| : ||:: | :::: ::|:
Db        226 -TGPDGESLTAKEAFEKAGISTGFFDLQPKEGLALVNGTAVGSGMASMVLFEANVQAVLA 284

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
               : |:|:  | | |:   |    :      : ||||:| |  :  :| |||: ::  ::|   |
Db        285 EVLSAIFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAIMEHILDGSSY-MKLAQKVHEMD 342

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQ 419
                 :||| |||||:||| :| :   |  ::  | |:  ||||:||     :  ||||||
Db        343 PLQKPKQDRYALRTSPQWLGPQIEVIRQATKSIEREINSVNDNPLIDVSRNKAIHGGNFQ 402

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGLDIH 478
              : : :||: ||||:| |||| | | :||:|   | ||||  | |: :|||:|  || :|
Db        403 GTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLTASSNPSLDYGFKGAEIA 462

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
              :|:| |||  :||||||: || ||  || |||| |||:|:|:|| |:| |:  :   |
Db        463 MASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLVGICQ 522

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRL------EQTTTY 592
              ||||| :| : ::      :  : :: | ||::        |    :    | |         || ||
Db        523 AVDLRHLEENLRQTVKNTVSQVAKKVLTTGINGE---LHPSRFCEKDLLKVVDREQVFTY 579

Qy        593 DLEP--RWHDAFSYATGTVVE-LLSSSPSANVTLTAVNAWKVASAEKAIS--LTREVRNR 647
                :|     :       :|:  ||:  :    :|::     |: : |: :     | :||
Db        580 VDDPCSATYPLMQRLRQVIVDHALSNGETEKNAVTSI-FQKIGAFEEELKAVLPKEVEAA 638

Qy        648 FWQTPSSQAPAHAYLSP-RTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAI 706
                 :  ||      :      |:  || ||||||| :    |:   |        :  |    ::: |:
Db        639 RAAYGNGTAPIPNRIKECRSYPLYRFVREELGTKLLTGEKVV-----SPGEEFDKVFTAM 693

Qy        707 KDGRINHVLVKML 719
               :|::  |:  |
Db        694 CEGKLIDPLMDCL 706
```

RESULT 10
S22991
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - soybean
C;Species: Glycine max (soybean)
C;Date: 06-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S22991
R;Frank, R.L.; Vodkin, L.O.
DNA Seq. 1, 335-346, 1991
A;Title: Sequence and structure of a phenylalanine ammonia-lyase gene from Glycine max
A;Reference number: S22991; MUID:92190550; PMID:1799682
A;Accession: S22991
A;Molecule type: DNA
A;Residues: 1-713 <FRA>
A;Cross-references: EMBL:X52953; NID:g18376; PIDN:CAA37129.1; PID:g18377
C;Genetics:
A;Gene: PAL1
A;Introns: 131/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;199-201/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;200/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.0%;  Score 1011;  DB 2;  Length 713;
  Best Local Similarity   36.9%;  Pred. No. 3.8e-59;
  Matches  270;  Conservative  120;  Mismatches  282;  Indels   60;  Gaps   19;

Qy         14 NGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVELSGYSL 73
              || |||    |   | |     |       |       |  | |: |: ::::     || | | :|
Db          5 NGHQNGSFCLST--AKGNNDPLNW--GAAAEAMKGSHLDEVKRMVAEYRKPVVRLGGETL 60

Qy         74 TVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRTE 131
              |:  |    |      | |: :  |  |  | :::   : |     ||||||| ::   ||:
Db         61 TIAQVAAVAGHDHGVAVELSESAREGVKASSEWVMNSMNNGTDSYGVTTGFGATSHRRTK 120

Qy        132 DAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVRL 191
               :||| ||     |:         |  |      :|||     | ||::|:|:| :|:| :|
Db        121 QGGALQKELIRFLNAGI-------FGNGTESSHTLPHTATRAAMLVRINTLLQGYSGIRF 173
```

```
Qy        192 VVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMF 251
Db        174 EILEAITKLLNNNVTPCLDLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPSG--EVLN 230

Qy        252 AREAISLFGL--EAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALTVEA 309
Db        231 AKEAFELASINSEFFELQPKEGLALVNGTAVGSGLASMVLFEANILAVLSEVLSAIFAEV 290

Qy        310 MVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVE----HEEEVKVKDDEGILR 365
Db        291 MQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSYMKAAKKLHEIDPLQKP-----K 344

Qy        366 QDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVSI 425
Db        345 QDRYALRTSPQWLGPLIEVIRFSTKSIEREINSVNDNPLIDVSRNKALHGGNFQGTPIGV 404

Qy        426 SMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGLDIHIAAYAS 484
Db        405 SMDNTRLALASIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCS 464

Qy        485 ELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLRA 544
Db        465 ELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTNEAIEILKLMSSTFLIALCQAIDLRH 524

Qy        545 MELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDAFSY 604
Db        525 LEENLKNSVKNTVSQVSKRILTTG--VNG-ELHPSRFCEKDLLKVVDREY------IFSY 575

Qy        605 ---ATGTVVELLSSSPSANVTLTAVNA-----------WKVASAEKAIS--LTREVRNRF 648
Db        576 IDDPCSATYPLMQKLRQVLVDHALVNAECEKDVNSSIFQKIAIFEEELKNLLPKEVEGAR 635

Qy        649 WQTPSSQAPAHAYLSP-RTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIK 707
Db        636 AAYESGKAAIPNKIQECRSYPLYKFVREELGT-----GLLTGEKVRSPGEEFDKLFTAMC 690

Qy        708 DGRINHVLVKML 719
Db        691 QGKIIDPLMECL 702
```

RESULT 11
S17444
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - alfalfa
C;Species: Medicago sativa (alfalfa)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S17444
R;Gowri, G.; Paiva, N.L.; Dixon, R.A.
Plant Mol. Biol. 17, 415-429, 1991
A;Title: Stress responses in alfalfa (Medicago sativa L.). 12. Sequence analysis of phen
nd developing plants.
A;Reference number: S17444; MUID:91355933; PMID:1715786
A;Accession: S17444
A;Molecule type: mRNA
A;Residues: 1-725 <GOW>
A;Cross-references: EMBL:X58180; NID:g19649; PIDN:CAA41169.1; PID:g19650
A;Experimental source: strain cv. Apollo
C;Genetics:
A;Gene: PAL1
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;212/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.0%;  Score 1009;  DB 2;  Length 725;
  Best Local Similarity   35.9%;  Pred. No. 5.3e-59;
  Matches  261;  Conservative  127;  Mismatches  241;  Indels   98;  Gaps   17;

Qy         40 GLDGHAAHQSQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRAR 99
Db         39 GVAAEAMKGSHLDEVKRMVAEYRKPVVRLGGETLTISQVAAIAAHDHGVQVDLSESARDG 98

Qy        100 VDKSVDFLKAQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFS 157
Db         99 VKASSEWVMESMNKGTDSYGVTTGFGATSHSRTKQGGALQKELIRFLNAGI-------FG 151

Qy        158 VGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISA 217
Db        152 NGTESNHTLPKTATRAAMLVRINTLLQGYSGIDFEILEAITKPLNKTVTPCLPLRGTITA 211

Qy        218 SGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLV 275
Db        212 SGDLVPLSYIAGLLTGRPNSKA---HGPSGEVLNAKEAFNLAGINAEFFELQPKEGLALV 268

Qy        276 NGTAVSASMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVA 335
Db        269 NGTAVGSGLASIVLFEANILAVLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAA 327

Qy        336 RNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE 395
Db        328 AIMEHILDGSSY-VKAAKKLHEIDPLQKPKQDRYALRTSPQWLGPLVEVIRFSTKSIERE 386

Qy        396 NNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNR 455
Db        387 INSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLALASIGKLMFAQFSELVNDFYNN 446

Qy        456 GLPSCL-AAEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALIS 514
Db        447 GLPSNLSASRNPSLDYGFKGAEIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLIS 506

Qy        515 ARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNAL 574
Db        507 ARKTNEAIEILQLMSSTFLIALCQAIDLRHLEENLK--------------------NSV 545

Qy        575 ALEVKKALNKRLEQTTTYDLEP------------RWHDAFSY------ATGTVVELL-- 613
Db        546 KNTVSQVAKKTLTMGVNGELHPSRFCEKDLLKVVDREH-VFAYIDDPCSATYPLSQKLRQ 604

Qy        614 ----------SSSPSANVTLTAVNAWKVASAEKAIS--LTREVRNRFWQTPSSQAPAHAY 661
Db        605 VLVDHALVNGESEKNFNTSIFQ----KIATFEEELKTLLPKEV--------ESARTAYES 652

Qy        662 LSP---------RTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRIN 712
Db        653 GNPTIPNKINGCRSYPLYKFVREELGTGLLTGENVI-----SPGEECDKLFSAMCQGKII 707

Qy        713 HVLVKML 719
Db        708 DPLLECL 714
```

RESULT 12
S04463
phenylalanine ammonia-lyase (EC 4.3.1.5) - parsley
C;Species: Petroselinum crispum (parsley)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-May-1999
C;Accession: S04463
R;Lois, R.; Dietrich, A.; Hahlbrock, K.; Schulz, W.
EMBO J. 8, 1641-1648, 1989
A;Title: A phenylalanine ammonia-lyase gene from parsley: structure, regulation and iden
A;Reference number: S04463; MUID:89356635; PMID:2767049
A;Accession: S04463
A;Molecule type: DNA
A;Residues: 1-716 <LOI>
A;Cross-references: EMBL:X15473
C;Genetics:
A;Introns: 136/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;202-204/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;203/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             27.9%;  Score 1006;  DB 2;  Length 716;
  Best Local Similarity   36.0%;  Pred. No. 8.2e-59;
  Matches  266;  Conservative  130;  Mismatches  275;  Indels   68;  Gaps   18;
```

```
Qy          9 ATTLANGFTNGSHAAPTKSAAGPTSALRRTP----GLDGHAAHQSQLEIVQELLSDPTDD 64
Db          7 ATT--NGHVNGN---------GMDFCMKTEDPLYWGIAAEAMTGSHLDEVKKMVAEYRKP 55

Qy         65 VVELSGYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTG 121
Db         56 VVKLGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDWVMDSMNKGTDSYGVTTG 115

Qy        122 FGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNS 181
Db        116 FGATSHRRTKQGGALQKELIRFLNAGI---------FGNGSDNTLPHSATRAAMLVRINT 166

Qy        182 LTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHV 241
Db        167 LLQGYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-- 224

Qy        242 LHEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
Db        225 -VGPTGVILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLA 283

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
Db        284 EVMSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEMD 341

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQ 419
Db        342 PLQKPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQ 401

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIH 478
Db        402 GTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIA 461

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
Db        462 MASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQ 521

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ---------- 588
Db        522 AIDLRHLEENLKSTVKNTVSSVAKRVLTMG--VNG-ELHPSRFCEKDLLRVVDREYIFAY 578

Qy        589 -----TTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS--LT 641
Db        579 IDDPCSATYPLMQKLRQ-------TLVEHALKNGDNERNLSTSIFQKIATFEDELKALLP 631

Qy        642 REVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVS 700
Db        632 KEVESARAALESGNPAIPNRIEECRSYPLYKFVRKELGT-----EYLTGEKVTSPGEEFE 686

Qy        701 RIYEAIKDGRINHVLVKML 719
Db        687 KVFIAMSKGEIIDPLLECL 705
```

RESULT 13
S25303
phenylalanine ammonia-lyase (EC 4.3.1.5) - garden pea
C;Species: Pisum sativum (garden pea)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Jun-2000
C;Accession: S25303
R;Kawamata, S.; Yamada, T.; Tanaka, Y.; Sriprasertsak, P.; Kato, H.; Ichinose, Y.; Kato,
Plant Mol. Biol. 20, 167-170, 1992
A;Title: Molecular cloning of phenylalanine ammonia-lyase cDNA from Pisum sativum.
A;Reference number: S25303; MUID:92385763; PMID:1515609
A;Accession: S25303
A;Molecule type: mRNA
A;Residues: 1-723 <KAW>
A;Cross-references: EMBL:D10001; NID:g217979; PIDN:BAA00885.1; PID:g217980
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;209-211/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;210/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             27.9%;  Score 1005.5;  DB 2;  Length 723;
  Best Local Similarity   35.8%;  Pred. No. 9e-59;
  Matches  270;  Conservative  135;  Mismatches  258;  Indels   91;  Gaps   23;

Qy          9 ATTLANGF-----TNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTD 63
Db          7 AITKNNGYESFCVTNAKNNNMKVNSADPLNW-----GVAAEAMKGSHLDEVKRMVEEYRK 61

Qy         64 DVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTG 121
Db         62 PVVRLGGETLTISQVAAIAAHDHGVKVELSESARAGVKASSDWVMESMNKGTDSYGVTTG 121

Qy        122 FGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNS 181
Db        122 FGATSHRRTKQGGALQKELIRFLNAGI-------FGNGTESSHTLPHTATRAAMLVRINT 174

Qy        182 LTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHV 241
Db        175 LLQGYSGIRFEILEAITKLINNNVTPCL-LRGTITASGDLVPLSYIAGLLTGRPNSKAH- 232

Qy        242 LHEGTE-KIMFAREAI-------SLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSH 293
Db        233 ---GTSGEILNAKEAFQSAEINDGFFELQ-----PKEGLALVNGTAVGSGLASIVLFEAN 284

Qy        294 MLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEE 353
Db        285 ILAVLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAK 342

Qy        354 EVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTA 413
Db        343 KLHEMDPLQKPKQDRYALRTSPQWLGPLIEVIRFSTKSIEREINSVNDNPLIDVSRNKAL 402

Qy        414 HGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHG 472
Db        403 HGGNFQGTPIGVSMDNTRLALASIGKLLFAQFSELVNDFYNNGLPSNLSASRNPSLDYGF 462

Qy        473 KGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASH 532
Db        463 KGSEIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTYEAIEILQLMSSTF 522

Qy        533 LYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTY 592
Db        523 LIALCQAVDLRHLEENLKNSVKNIVSQVAKRTLTTG--VNG-ELHPSRFCEKDLLRVV-- 577

Qy        593 DLEPRWHDAFSY------ATGTVVE----------LLSSSPSANVTLTAVNAWKVASAEK 636
Db        578 ---DREH-VFAYIDDPCSATYPLMQKLRQVLVDHALVNGESEKNLNTSIFQ--KIATFED 631

Qy        637 AIS--LTREVRNRFWQTPSSQAPAHAYLSP---------RTRVLYSFVREELGVQARRGD 685
Db        632 ELKTLLPKEV--------ESTRAAYESGNPTVPNKINGCRSYPLYRFVRQELGTGLLTGE 683

Qy        686 VFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719
Db        684 KVI-----SPGEECDKLFTAICQGKIIDPLLQCL 712
```

RESULT 14
T03663
phenylalanine ammonia-lyase (EC 4.3.1.5) - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Jul-1999
C;Accession: T03663
R;Pellegrini, L.; Rohfritsch, O.; Fritig, B.; Legrand, M.
Plant Physiol. 106, 877-886, 1994
A;Title: Phenylalanine ammonia-lyase in tobacco. Molecular cloning and gene expression
A;Reference number: Z14995; MUID:95125127; PMID:7824656
A;Accession: T03663
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-712 <PEL>
A;Cross-references: EMBL:X78269; NID:g633596; PIDN:CAA55075.1; PID:g633597
A;Experimental source: cultivar Samsun NN; tissue-type leaves after infection by tobac

```
C;Genetics:
A;Gene: PAL
C;Function:
A;Description: catalyzes the deamination of L-phenylalanine to trans-cinnamic acid
A;Note: strongly induced during the hypersensitive reaction to TMV or to a fungal elicit
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;198-200/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;199/Modified site: dehydroalanine (Ser) #status predicted

  Query Match             27.8%;  Score 1005;  DB 2;  Length 712;
  Best Local Similarity   37.4%;  Pred. No. 9.5e-59;
  Matches  260;  Conservative  131;  Mismatches  253;  Indels   52;  Gaps   17;

Qy         49 SQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARK--GRRVRVQNDDEIRARVDKSVDF 106
Db         33 SHLDEVKKMVAEFRKPVVKLGGETLTVAQVAAIAAKDNAKTVKVELSEGARAGVKASSDW 92

Qy        107 LKAQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE- 163
Db         93 VMDSMSKGTDSYGVTTGFGATSHRRTKNGGALQKELIRFLNAGV---------FGNGTES 143

Qy        164 -NTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLS 222
Db        144 CHTLPQSGTRAAMLVRINTLLQGYSGIRFEILEAITKLLNHNVTPCLPLRGTITASGDLV 203

Qy        223 PLSYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAV 280
Db        204 PLSYIAGLLTGRPNSKA-IGPNG--ETLNAEEAFRVAGVNSGFFELQPKEGLALVNGTAV 260

Qy        281 SASMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRT 340
Db        261 GSGLASMVLFDANILAVFSEVLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEH 319

Qy        341 LLSGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTT 400
Db        320 ILDGSSY-VKAPQKLHETDPLQKPKQDRYALRTSPQWLGPQIEVIRSATKMIEREINSVN 378

Qy        401 DNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSC 460
Db        379 DNPLIDVSRNKALHGGNFQGTPIGVSMDNARLALASIGKLMFAQFSELVNDYYNNGLPSN 438

Qy        461 L-AAEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTA 519
Db        439 LTAGRNPSLDYGFKGSEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISARKTA 498

Qy        520 EANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVK 579
Db        499 EAVDILKLMSSTYLVALCQAIDLRHLEENLRNAVKNTVSQVAKRTLTMGANGE---LHPS 555

Qy        580 KALNKRLEQTTTYDLEPRWHDAFSYATGTVVELL-----------SSSPSANVTLTAVN 627
Db        556 RFCEKDLLRVVDREYVFRYADDACSANYPLMQKLRQVLVDHALENGENEKNANSSIFQ-- 613

Qy        628 AWKVASAE---KAISLTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARR 683
Db        614 --KILAFEGELKAV-LPKEVESARISLENGNPAIANRIKECRSYPLYRFVREELG----- 665

Qy        684 GDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719
Db        666 AELLTGEKVRSPGEECDKVFTAMCNGQIIDSLLECL 701

RESULT 15
S48725
phenylalanine ammonia-lyase (EC 4.3.1.5) 2 - parsley
C;Species: Petroselinum crispum (parsley)
C;Date: 28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48725
R;Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
Eur. J. Biochem. 225, 491-499, 1994
A;Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase isc
A;Reference number: S48725; MUID:95010141; PMID:7925471
A;Accession: S48725
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-716 <APP>
A;Cross-references: EMBL:X81158; NID:g534892; PIDN:CAA57056.1; PID:g534893
C;Genetics:
A;Gene: PAL2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;202-204/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;203/Modified site: dehydroalanine (Ser) #status predicted

  Query Match             27.8%;  Score 1005;  DB 2;  Length 716;
  Best Local Similarity   35.8%;  Pred. No. 9.6e-59;
  Matches  264;  Conservative  133;  Mismatches  271;  Indels   70;  Gaps   19;

Qy         12 LANGFTNGSHAAPTKSAAGPTSALRRTP----GLDGHAAHQSQLEIVQELLSDPTDDVVE 67
Db          8 ITNGHVNGN---------GMDFCMKTEDPLYWGIAAEAMTGSHLDEVKKMVAEYRKPVVK 58

Qy         68 LSGYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGG 124
Db         59 LGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDWVMDSMNKGTDSYGVTTGFGA 118

Qy        125 SADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTR 184
Db        119 TSHRRTKQGGALQKELIRFLNAGI---------FGNGSDNTLPHSATRAAMLVRINTLLQ 169

Qy        185 GHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHE 244
Db        170 GYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA---VG 226

Qy        245 GTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQAL 302
Db        227 PTGVILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEVM 286

Qy        303 TALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEG 362
Db        287 SAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEMDPLQ 344

Qy        363 ILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASA 422
Db        345 KPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGTP 404

Qy        423 VSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAA 481
Db        405 IGMSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMAS 464

Qy        482 YASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVD 541
Db        465 YCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAID 524

Qy        542 LRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ------------ 588
Db        525 LRHLEENLKSTVKNTVSSVAKRVLTMG--VNG-ELHPSRFCEKDLLRFVDREYIFAYIDD 581

Qy        589 --TTTYDLEPRWHDAFSYATGTVVE--LLSSSPSANVTLTAVNAWKVASAEKAIS--LTR 642
Db        582 PCSATYPLMQKLRQ-------TLVEHALKNGDNERNMNTSIFQ--KIATFEDELKALLPK 632

Qy        643 EVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSR 701
Db        633 EVESARAALESGNPAIPNRIEECRSYPLYKFVRKELGI-----EYLTGEKVTSPGEEFDK 687

Qy        702 IYEAIKDGRINHVLVKML 719
Db        688 VFIAMSKGEIIDPLLECL 705

Search completed: September  9, 2004, 10:08:45
Job time : 32.8797 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:         September  9, 2004, 09:50:23 ; Search time 16.4315 Seconds
                                           (without alignments)
                                           2281.620 Million cell updates/sec

Title:          US-09-939-408A-13
Perfect score:  3610
Sequence:       1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2614	72.4	716	1	PALY_RHOTO	P11544 rhodosporid
2	2499.5	69.2	713	1	PALY_RHORB	P10248 rhodotorula
3	1346.5	37.3	740	1	PALY_AMAMU	O93967 amanita mus
4	1247.5	34.6	724	1	PALY_USTMA	Q96v77 ustilago ma
5	1035.5	28.7	725	1	PAL1_ARATH	P35510 arabidopsis
6	1024	28.4	717	1	PAL1_PRUAV	O64963 prunus aviu
7	1020	28.3	707	1	PAL4_ARATH	Q9ss45 arabidopsis
8	1019	28.2	725	1	PALY_TRISU	P45734 trifolium s
9	1017.5	28.2	718	1	PAL3_PETCR	P45729 petroselinu
10	1016.5	28.2	715	1	PAL1_TOBAC	P25872 nicotiana t
11	1016	28.1	715	1	PALY_POPTR	P45730 populus tri
12	1015.5	28.1	717	1	PAL2_ARATH	P45724 arabidopsis
13	1013.5	28.1	713	1	PALY_DIGLA	O23924 digitalis l
14	1012	28.0	714	1	PALY_CAMSI	P45726 camellia si
15	1011	28.0	713	1	PAL1_SOYBN	P27991 glycine max
16	1009	28.0	725	1	PALY_MEDSA	P27990 medicago sa
17	1006	27.9	716	1	PAL1_PETCR	P24481 petroselinu
18	1005.5	27.9	723	1	PAL1_PEA	Q01861 pisum sativ
19	1005	27.8	712	1	PAL3_TOBAC	P45733 nicotiana t
20	1005	27.8	716	1	PAL2_PETCR	P45728 petroselinu
21	1004	27.8	715	1	PALY_STYHU	P45732 stylosanthe
22	1003.5	27.8	712	1	PAL2_PHAVU	P19142 phaseolus v
23	998	27.6	701	1	PAL1_ORYSA	P14717 oryza sativ
24	998	27.6	705	1	PAL2_LITER	O49836 lithospermu
25	997.5	27.6	721	1	PAL5_LYCES	P26600 lycopersico
26	997	27.6	720	1	PAL1_SOLTU	P31425 solanum tub
27	995.5	27.6	704	1	PAL1_LYCES	P35511 lycopersico
28	995.5	27.6	724	1	PAL2_PEA	Q04593 pisum sativ
29	995	27.6	710	1	PAL1_LITER	O49835 lithospermu
30	993.5	27.5	708	1	PAL2_IPOBA	Q42858 ipomoea bat
31	993	27.5	712	1	PAL2_TOBAC	P35513 nicotiana t
32	991	27.5	718	1	PAL2_CICAR	Q9smk9 cicer ariet
33	987.5	27.4	667	1	PALY_HELAN	O04058 helianthus
34	983	27.2	708	1	PAL1_DAUCA	O23865 daucus caro
35	982.5	27.2	710	1	PAL2_ORYSA	P53443 oryza sativ
36	980	27.1	710	1	PAL1_RUBID	Q9m568 rubus idaeu
37	979.5	27.1	722	1	PALY_CITLI	Q42667 citrus limo
38	973.5	27.0	700	1	PALY_WHEAT	Q43210 triticum ae
39	973.5	27.0	754	1	PALY_PINTA	P52777 pinus taeda
40	966	26.8	590	1	PAL2_SOLTU	P31426 solanum tub
41	964.5	26.7	695	1	PAL3_ARATH	P45725 arabidopsis
42	954.5	26.4	707	1	PAL1_IPOBA	P14166 ipomoea bat
43	953.5	26.4	710	1	PAL2_POPKI	Q43052 populus kit
44	947	26.2	682	1	PAL1_POPKI	P45731 populus kit
45	946.5	26.2	710	1	PAL3_PHAVU	P19143 phaseolus v

ALIGNMENTS

```
RESULT 1
PALY_RHOTO
ID   PALY_RHOTO     STANDARD;      PRT;   716 AA.
AC   P11544;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Rhodosporidium toruloides (Yeast) (Rhodotorula gracilis).
OC   Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
OC   Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
OX   NCBI_TaxID=5286;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 14;
RX   MEDLINE=92127061; PubMed=1773059;
RA   Rasmussen O.F., Orum H.;
RT   "Analysis of the gene for phenylalanine ammonia-lyase from
RT   Rhodosporidium toruloides.";
RL   DNA Seq. 1:207-211(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 0559;
RX   MEDLINE=88112870; PubMed=2828184;
RA   Anson J.G., Gilbert H.J., Oram J.D., Minton N.P.;
RT   "Complete nucleotide sequence of the Rhodosporidium toruloides gene
RT   coding for phenylalanine ammonia-lyase.";
RL   Gene 58:189-199(1987).
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X51513; CAA35886.1; -.
DR   EMBL; M18261; AAA33883.1; -.
DR   EMBL; X12702; CAA31209.1; -.
DR   PIR; A29607; A29607.
DR   PIR; A56628; A56628.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
```

```
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    211    213       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     212    212       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
FT   CONFLICT      4     37       SLDSISHSFANGVASAKQAVNGASTNLAVAGSHL ->
FT                                RPTSQSQARTC (IN REF. 2).
SQ   SEQUENCE   716 AA;  76879 MW;  0C1DF61769A4E5E6 CRC64;

  Query Match             72.4%;  Score 2614;  DB 1;  Length 716;
  Best Local Similarity   72.0%;  Pred. No. 2.4e-167;
  Matches  522;  Conservative   86;  Mismatches  103;  Indels   14;  Gaps    5;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60
              |||||||:: : |||        |  |  |    |      |      :|::||:::|:
Db          1 MAPSLDSISHSFANG------VASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVTT 120
              |||  :|| |||| : ||| ||||| |||:: ||||:::||||:||::||  ||||||
Db         55 PTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVTT 114

Qy        121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
              ||||||||||||:||||||:|||||:|||||| |:|  || :|||||:|||||||||| |||
Db        115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVN 174

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              |||||||||||||||||||||| |||||||||:||||||||||||||| ||:|||| |||
Db        175 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKVH 234

Qy        241 VLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |:||| |||::||||::|| || ||||||||||||||||||||||:|||:||||||||
Db        235 VVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 294

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDD 360
              :|||:||||||  ||| ||:||| |||| |:||| ||| || || ||| |||||||||
Db        295 SLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDD 354

Qy        361 EGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNFQ 419
              |||||||||||||||||:||||| |::||::  |::|    :||||||:|||| : |||||
Db        355 EGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQ 414

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIHI 479
              |:||: :||||| || ||||||| ||:||| ||||||||||||||:|| |||||
Db        415 AAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAA 474

Qy        480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQA 539
              ||| ||||||||||| |||||| |||||||||||||| |:|||||||||:|||| |||
Db        475 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQA 534

Qy        540 VDLRAMELDFKKQFDPLLPTLLQQHLG---TGLDV-NALALEVKKALNKRLEQTTTYDLE 595
              :||||:| :|||||  | : :|: || |    ||  :: :  |   :| | |  ||||| :|||
Db        535 IDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLV 594

Qy        596 PRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSSQ 655
              ||||||||:| |||||:|||:     :::| ||||||||:|| ||||||:||  ||    |:
Db        595 PRWHDAFSFAAGTVVEVLSST---SLSLAAVNAWKVAAAESAISLTRQVRETFWSAASTS 651

Qy        656 APAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVL 715
              :|| :|||||::||:|||||||||:|||||||:| |: ||||||:|||||| ||||:||
Db        652 SPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVL 711

Qy        716 VKMLA 720
              :||||
Db        712 LKMLA 716

RESULT 2
PALY_RHORB
ID   PALY_RHORB     STANDARD;      PRT;   713 AA.
AC   P10248;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Rhodotorula rubra (Yeast) (Rhodotorula mucilaginosa).
OC   Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
OC   Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
OC   Rhodotorula.
OX   NCBI_TaxID=5537;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL Y-15597;
RX   MEDLINE=89083582; PubMed=3205749;
RA   Filpula D., Strausberg R.L., Vaslet C.A., Sykes A., Levy A.;
RT   "Nucleotide sequence of gene for phenylalanine ammonia-lyase from
RT   Rhodotorula rubra.";
RL   Nucleic Acids Res. 16:11381-11381(1988).
RN   [2]
RP   DETAILS.
RC   STRAIN=NRRL Y-15597;
RX   MEDLINE=89083583; PubMed=3205750;
RA   Filpula D., Strausberg R.L., Vaslet C.A., Sykes A., Levy A.;
RT   "cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes
RT   reveal genomic intron deletions.";
RL   Nucleic Acids Res. 16:11382-11382(1988).
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X13094; CAA31486.1; -.
DR   PIR; S01999; S01999.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    217    219       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     218    218       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   713 AA;  76000 MW;  6EB8317CBB037D88 CRC64;

  Query Match             69.2%;  Score 2499.5;  DB 1;  Length 713;
  Best Local Similarity   69.6%;  Pred. No. 1.1e-159;
  Matches  511;  Conservative   83;  Mismatches  105;  Indels   35;  Gaps    9;

Qy          1 MAPSLDSLATTLANGFTNGSHAAPT-------KSAAGPTSALRRTPGLDGHAAHQSQLEI 53
              ||||:||:||::||  :|| |||             |  ||   |  |  |          :||:|
Db          1 MAPSVDSIATSVANSLSNGLHAAAAANGGDVHKKTAGAGSLLPTT--------ETTQLDI 52

Qy         54 VQELLSDP-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQ 112
              |: :|:|     | ::| ||:||: |||||||:|| |:| :      || ::| ||:||: ||
Db         53 VERILADAGATDQIKLDGYTLTLGDVVGAARRGRSVKVADSPHIREKIDASVEFLRTQLD 112

Qy        113 NSVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVR 172
```

```
Db        113 NSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSMDGFALGRGLENSLPLEVVR 172

Qy        173 GAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAIT 232

Db        173 GAMTIRVNSLTRGHSAVRIVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAASIT 232

Qy        233 GHPDVKVHVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDS 292

Db        233 GHPDSKVHV--DG--KIMSAQEAIALKGLQPVVLGPKEGLGLVNGTAVSASMATLALTDA 288

Qy        293 HMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHE 352

Db        289 HVLSLLAQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSKYAVHHE 348

Qy        353 EVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQ 411

Db        349 TEVKVKDDEGILRQDRYPLRCSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDLENKM 408

Qy        412 TAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYH 471

Db        409 THHGGAFMASSVGNTMEKTRLAVALMGKVSFTQLTEMLNAGMNRALPSCLAAEDPSLSYH 468

Qy        472 GKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLAS 531

Db        469 CKGLDIAAAAYTSELGHLANPVSTHVQPAEMGNQAINSLALISARRTAEANDVLSLLLAT 528

Qy        532 HLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALAL-----EVKKALNKRL 586

Db        529 HLYCVLQAVDLRAMEFEHTKAFEPMVTELLKQHFG------ALATAEVEDKVRKSIYKRL 582

Qy        587 EQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRN 646

Db        583 QQNNSYDLEQRWHDTFSVATGAVVEALAGQ---EVSLASLNAWKVACAEKAIALTRSVRD 639

Qy        647 RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAI 706

Db        640 SFWAAPSSSSPALKYLSPRTRVLYSFVREEVGVKARRGDVYLGKQEVTIGTNVSRIYEAI 699

Qy        707 KDGRINHVLVKMLA 720

Db        700 KSGCIAPVLVKMMA 713
```

RESULT 3

```
PALY_AMAMU
ID   PALY_AMAMU     STANDARD;      PRT;   740 AA.
AC   O93967;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Amanita muscaria (Fly agaric).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Agaricales; Amanitaceae; Amanita.
OX   NCBI_TaxID=41956;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nehls U., Mikolajewski S., Ecke M., Hampp R.;
RT   "Sugar and nitrogen-dependent regulation of an Amanita muscaria
RT   phenylalanine ammonium lyase gene.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ser-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ010143; CAA09013.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    224    226       5-imidazolinone (Ser-Gly)
FT                                (By similarity).
FT   MOD_RES     225    225       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   740 AA;  80156 MW;  B055CFF3D8B7BCEE CRC64;

  Query Match            37.3%;  Score 1346.5;  DB 1;  Length 740;
  Best Local Similarity  43.0%;  Pred. No. 2e-82;
  Matches  317;  Conservative  131;  Mismatches  238;  Indels   51;  Gaps   15;

Qy         10 TTLANGF--TNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQ-LEIVQELLSDPTDDVV 66

Db         25 TTPVNGFKATALSKASRTMT---KTSAL------------SQFLEAYRELEGYKNGRAI 68

Qy         67 ELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFL--KAQLQNSVYGVTTGFGG 124

Db         69 KVDGQTLSIAAVAAAARYNAAVELDESPLVKERVRKSQLAIANKVSTGASVYGLSTGFGG 128

Qy        125 SADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSV--GRGLENT-LPLEVVRGAMVIRVNS 181

Db        129 SADTRTDKPMLLGFALLQHQHVGILPTSTEPLDVLPLQDANNTSMPEAWIRGAILIRMNS 188

Qy        182 LTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHV 241

Db        189 LIRGHSGIRWELIEKMRELLAANVIPVVPLRGSISSGDLSPLSYIAGTIIGNPSIKVY- 247

Qy        242 LH----EGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSL 297

Db        248 -HGPSKSGIRQIGSSKDVLALHNIEPPLESKEPLGILNGTAFSASVAALALNEAIHLVL 306

Qy        298 LSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKV 357

Db        307 LAQVCTAMGTEALIGTRASHAPFIHATARPHPGQVECAENIWNLLDGSKLAQLEEHEVRL 366

Qy        358 KDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENN--TTTDNPLLDVENKQTAHG 415

Db        367 EDDKYTLRQDRYPLRTSPQFLGPQIEDIISAFQTVTQECNYLPATDNPLIDGETGESHHG 426

Qy        416 GNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGL 475

Db        427 GNFQAMAVTNAMEKTRLALHHVGKLLFSQSTELVNPAMNRGLPPSVAATDPSLNYHAKGL 486

Qy        476 DIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYC 535

Db        487 DIATAAYVAE----ATPGPTHIQSAEMHNQAVNSLALISARATITSLEVLTSLIASYLYI 542

Qy        536 TLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGL---DVNALALEVKKALNKRLEQTTTY 592

Db        543 LCQALDLRALQREFLPGLDIIIREELRSSFGSFLSSEQMEKLQQNLTSAFEDHLDKTTTM 602

Qy        593 DLEPRWHDAFSYATGTVVELLSSS-----PSANVTLTAVNAWKVASAEKAISLTREVRNR 647

Db        603 DNTDRMTTMAATSSSVLLQFFTDSGASVPPSSCDLLSSVSSFQSSVATRSSVLMDDLRKE 662

Qy        648 FWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVF------VGVQQETIGSNVSR 701
```

```
Db        663 YIFGDRGPTPASQYIG-KTRPVYQFIRTTIGVRKHGSENYNKFYNGLGVEDVTIGQNISR 721

Qy        702 IYEAIKDGRINHVLVKM 718
              |||:|:||::  ::| :
Db        722 IYESIRDGKMQSIIVSL 738


RESULT 4
PALY_USTMA
ID   PALY_USTMA     STANDARD;      PRT;   724 AA.
AC   Q96V77;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL1.
OS   Ustilago maydis (Smut fungus).
OC   Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC   Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX   NCBI_TaxID=5270;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim S.-H., Macdonald K., Virmani D., Wake K., Kronstad J.W.,
RA   Ellis B.E.;
RT   "Cloning and disruption of a phenylalanine ammonia-lyase gene from
RT   Ustilago maydis.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF306551; AAL09388.1; -.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    205    207       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     206    206       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   724 AA;  79277 MW;  02A73C97C34CD1AB CRC64;

  Query Match             34.6%;  Score 1247.5;  DB 1;  Length 724;
  Best Local Similarity   40.8%;  Pred. No. 8.1e-76;
  Matches  302;  Conservative  130;  Mismatches  243;  Indels   65;  Gaps   18;

Qy         23 APTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTD-----------DVVELSGY 71
              |||     |   |  |  ||                   || |:|        : | : ||
Db          2 APTADVLPPVEASTR-PG----------------LLVQPSDTKLRKASSFRTEQVVIDGY 44

Qy         72 SLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQ--NSVYGVTTGFGGSADTR 129
              :| :: :| :||  |    | :     |  | |:| ||  |  |:|    |:||: ||||||||:|
Db         45 NLKIQGLVASARYGHVTRPRPSAETRKRIDDSVQSLIAKLDGGESIYGINTGFGGSADSR 104

Qy        130 TEDAVSLQKALIEHQLCGVTPTSXSSFSVGR---------GLENT-----LPLEVVRGAMV 176
              | :  :|| ||:: | ||| |   |:|  |             | :|     :|   ||||:|
Db        105 TANTRALQLALLQMQQCGVLPVP-STFPTGEPSSAPFALPLTDTESSLIMPEAWVRGAIV 163

Qy        177 IRVNSLTRGHSAVRLVVLEALTN-FLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHP 235
              :|::|| |||| ||  ||: :    || : :||:||:| |||||||||||||:|||: |
Db        164 VRLSSLMRGHSGVRWEVLDKMQKLFLQNNVTPVVPVRSSISASGDLSPLSYVAGALAGQR 223

Qy        236 DVKVHVLHEGTEKI-MFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHM 294
               :  |     ::: : | ||  :  :   |    ||| |||:||||  |||:|  |: :::
Db        224 GIYCFVTDGRGQRVKVTADEACRMHKITPVQYEPKEALGLLNGTAFSASVAGLATYEAEN 283

Qy        295 LSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFA--VEHE 352
              |: |:|  ||: |||: |    |||||||:: ||||||:: |: || |:|   |   | :|:|
Db        284 LASLTQLTTAMAVEALKGTDASFAPFIHEIARPHPGQIKSAKFIRAHLSGSRLAEHLENE 343

Qy        353 EEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDV--ENK 410
               : |     :| |  ||||||  |:|:  |::|| :||: :|  ::   | |:||||||:: |
Db        344 KHVLFSEDNGTLRQDRYTLQTASQWVGPGLEDIENAKRSVDFEINSTTDNPMIDPYDGDG 403

Qy        411 QTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAA-EDPSLN 469
               :  ||||||| |:: ::|| ||||  :||:  |  | |||:| ||||||: ||:  | |||
Db        404 RIHHGGNFQAMAMTNAVEKIRLALCAMGKMTFQQMTELVNPAMNRGLPANLASTPDLSLN 463

Qy        470 YHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLL 529
              :| ||::| :|: |||  | |||:| || ||| ||| ||||||| |:| :| : ||::
Db        464 FHAKGINIALASVTSELMFLGNPVSTHVQSAEMANQAFNSLALISGRQTLQAIECLSMIQ 523

Qy        530 ASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLD---VNALALEVKKALNKRL 586
              | ||  ||:|:||::     :|    |:    |   |   |  |  :|    :| :| |::|
Db        524 AWSLYLLCQALDIRALQYKVAEQLPTLILASLHSHFGEWMDETKQQEIAAQVLKSMSKRL 583

Qy        587 EQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVT--LTAVNAWKVASAEKAISLTREV 644
              ::|::  ||  |  : :  |: :|  |  ||       |  :  |:          : |:|
Db        584 DETSSKDLRDRLVETYQDASSVLVRYFSELPSGGGADPLRNIVKWRATGVADTEKIYRQV 643

Qy        645 RNRFWQTPSSQAPAHA-YLSPRTRVLYSFVREELGVQARRGD---VFVG--VQQETIGSN 698
               |  |  :   || :|  :|:  | |||: |||    :     ||    |  | |
Db        644 TIEFLDNPYA---CHASHLLGKTKRAYEFVRKTLGVPMHGKENLNEFKGEFEQWNTTGGY 700

Qy        699 VSRIYEAIKDGRINHVLVKM 718
              || || :|:|| : ::| ::
Db        701 VSVIYASIRDGELYNMLSEL 720


RESULT 5
PAL1_ARATH
ID   PAL1_ARATH     STANDARD;      PRT;   725 AA.
AC   P35510; Q9ZQD6;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN   PAL1 OR AT2G37040 OR T1J8.22.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=95195160; PubMed=7888622;
RA   Wanner L.A., Li G., Ware D., Somssich I.E., Davis K.R.;
RT   "The phenylalanine ammonia-lyase gene family in Arabidopsis
RT   thaliana.";
RL   Plant Mol. Biol. 27:327-338(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
```

```
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [3]
RP   SEQUENCE OF 1-240 FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=93005677; PubMed=2152131;
RA   Ohl S., Hedrick S.A., Chory J., Lamb C.J.;
RT   "Functional properties of a phenylalanine ammonia-lyase promoter from
RT   Arabidopsis.";
RL   Plant Cell 2:837-848(1990).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L33677; AAC18870.1; -.
DR   EMBL; AC006922; AAM15324.1; -.
DR   EMBL; X62747; CAA44609.1; -.
DR   PIR; S52990; S52990.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    211    213       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     212    212       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
FT   CONFLICT    329    329       V -> I (IN REF. 2).
FT   CONFLICT    426    426       R -> A (IN REF. 2).
FT   CONFLICT    612    612       V -> I (IN REF. 2).
SQ   SEQUENCE   725 AA;  78782 MW;  5D65F102D50A2946 CRC64;

  Query Match             28.7%;  Score 1035.5;  DB 1;  Length 725;
  Best Local Similarity   36.4%;  Pred. No. 1.3e-61;
  Matches  271;  Conservative  138;  Mismatches  259;  Indels   77;  Gaps   21;

Qy         18 NGSHAAPTKSAAGPTSAL--------------RRTPGLDGHAAHQ---SQLEIVQELLSD 60
              ||:|      ||  |   |:                |   | ||  |    | |: |: ::::
Db          4 NGAH----KSNGGGVDAMLCGGDIKTKNMVINAEDPLNWGAAAEQMKGSHLDEVKRMVAE 59

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGV 118
                  || | | :|:  |    :  | |:|:  :  || |: | |:  :         |||
Db         60 FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARAGVNASSDWVMESMNKGTDSYGV 119

Qy        119 TTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIR 178
              ||||| ::   ||:: |:||| ||        |:          |     :|||    | ||::|
Db        120 TTGFGATSHRRTKNGVALQKELIRFLNAGI-------FGSTKETSHTLPHSATRAAMLVR 172

Qy        179 VNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVK 238
               :|:| :| | :|   :|||:|:|||: ||| :||||:|:||||| ||||||| :|| |: |

Db        173 INTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTITASGDLVPLSYIAGLLTGRPNSK 232

Qy        239 VHVLHEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHML 295
                     |   : :  | ||  |  |: :      | |||||  ||||||  : ||:: | ::::|
Db        233 A----TGPNGEALTAEEAFKLAGISSGFFDLQPKEGLALVNGTAVGSGMASMVLFETNVL 288

Qy        296 SLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEV 355
              |:|:: |:|:  | | |:   |     :     : ||||:| |   :  :| |||: ::  :::
Db        289 SVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAVMEHILDGSSY-MKLAQKL 346

Qy        356 KVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHG 415
                  |    :|||  |||||:|||  :|  : :|   ::    |  |:  ||||:||     :   ||
Db        347 HEMDPLQKPKQDRYALRTSPQWLGPQIEVIRYATKSIEREINSVNDNPLIDVSRNKAIHG 406

Qy        416 GNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKG 474
              |||| :  : :||: ||||:  ||||  |  |  :||:|    |  ||||  |  |: :|||:|   ||
Db        407 GNFQGTPIGVSMDNTRLAIRAIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKG 466

Qy        475 LDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLY 534
               :|  :|:|  |||  :|||||:  || ||   || ||||  |||:|:|:||  |:|  |:  :  |
Db        467 AEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLV 526

Qy        535 CTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRL------EQ 588
               ||||||  :|   :  :    :     :  : ::  |  ||  ||    |    |   |   ||
Db        527 AICQAVDLRHLEENLRQTVKNTVSQVAKKVLTTG--VNG-ELHPSRFCEKDLLKVVDREQ 583

Qy        589 TTTYDLEPRWHDAFSYATGTVVE----------LLSSSPSANVTLTAVNAWKVASAEKAI 638
               ||  :|        |  :::               |::     |    :    :  |: :  |: :
Db        584 VYTYADDP------CSATYPLIQKLRQVIVDHALVNGESEKNAVTSIFH--KIGAFEEEL 635

Qy        639 S--LTREVR--NRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQET 694
                  | :||       :     :|   |  :           |:   ||  ||||||     ::   |  :  :
Db        636 KAVLPKEVEAARAAYDNGTSAIP-NRIKECRSYPLYRFVREELGT-----ELLTGEKVTS 689

Qy        695 IGSNVSRIYEAIKDGRINHVLVKML 719
                |      :::  ||  :|:|     ::: |
Db        690 PGEEFDKVFTAICEGKIIDPMMECL 714

RESULT 6
PAL1_PRUAV
ID   PAL1_PRUAV     STANDARD;      PRT;   717 AA.
AC   O64963;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN   PAL1.
OS   Prunus avium (Cherry).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX   NCBI_TaxID=42229;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Summit;
RA   Wiersma P.A., Wu Z.;
RT   "A full-length cDNA for phenylalanine ammonia-lyase cloned from ripe
RT   Sweet Cherry fruit (Prunus avium).";
RL   (In) Plant Gene Register PGR98-184.
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF036948; AAC78457.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    203    205       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     204    204       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   717 AA;  77999 MW;  B84DF90FA0BF60B3 CRC64;

  Query Match             28.4%;  Score 1024;  DB 1;  Length 717;
  Best Local Similarity   36.2%;  Pred. No. 7.4e-61;
  Matches  262;  Conservative  133;  Mismatches  269;  Indels   60;  Gaps   17;

Qy         14 NGFTNGSHAAPTKSAAGPTSALRRTP---GLDGHAAHQSQLEIVQELLSDPTDDVVELSG 70
              ||  ||          |    |      :::  |    |:            | |:  |: ::::     ||:| |
Db          9 NGHKNG-------SVELPELCIKKDPLNWGVAAETLKGSHLDEVKRMVAEYRKPVVKLGG 61

Qy         71 YSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADT 128
              |||:  |    |    |:|:  :  ||  |   | |||::    :          |||||||  ::
Db         62 ESLTISQVAAIATHDSGVKVELSESARAGVKASSDWVMDSMSKGTDSYGVTTGFGATSHR 121

Qy        129 RTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGHSA 188
              ||:   :|||  ||        ||   ::    |              :|||  :    | ||::|:|:|  :|:|
Db        122 RTKQGAALQKELIRFLNAGVFGSTKES-------GHTLPHQATRAAMLVRINTLLQGYSG 174

Qy        189 VRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEK 248
              :|   :||  :|  |||:  :||  :||||:|:|||||  |||||||  :||  |:  |   |    :|    :
Db        175 IRFEILEVITKFLNNNVTPCLPLRGTITASGDLVPLSYIAGMLTGRPNSKA-VGPDG--Q 231

Qy        249 IMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALT 306
              :  |  ||       |:  :         |||||  |||||||  : :|:   |  |:::|:||: |:|:
Db        232 TLSAAEAFEFVGINSGFFELQPKEGLALVNGTAVGSGLASTVLFDTNILALLSEILSAIF 291

Qy        307 VEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVE----HEEEVKVKDDEG 362
              |  |  |:      |        :       :  ||||:|  |     :   :|  |||:           ||::     |
Db        292 AEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSYVKAAKKLHEQDPLQKP--- 347

Qy        363 ILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASA 422
              :|||  |||||:|||  :|   :  ::   ::    |  ::     ||||:||    :    ||||||   :
Db        348 --KQDRYALRTSPQWLGPQIEVIRYSTKSIEREIDSVNDNPLIDVSRNKALHGGNFQGTP 405

Qy        423 VSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAA 481
              :  :||:   |||||:|  ||||  |   |  :||:|      |   |:|||:|    ||:  |:  |:|:
Db        406 IGVSMDNTRLAIASIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMAS 465

Qy        482 YASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVD 541
              |  |||   ||||||   ||  ||       ||  ||||  |||:|:|||  |:|   |:  ::   |    ||:|
Db        466 YCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSSTFLVALCQAID 525

Qy        542 LRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALE-VKKALNKRLEQ----------T 589
              ||  :|  :  :          :   :  ::  | ||::         :|  |  |  :::                :
Db        526 LRHLEENLRNTVKNTVSQVAKRTLTTGVNGELHPSRFCEKDLLKVVDREYVFAYIDDPCS 585

Qy        590 TTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEK-AISLTREVRNRF 648
              ||  |    :                |    |::   :          |::    ||   |:   :   |  :||   :
Db        586 ATYPLMQKLRQVL------VEHALTNGENEKNASTSIFQKIVAFEEELKVLLPKEVDSAR 639

Qy        649 WQTPSSQAPAHAYLSP-RTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIK 707
              |      |       ::   |:   ||  |||||||          :     |   :   :  |      :::  ||
Db        640 AALDSGSAGVPNRITECRSYPLYKFVREELG-----AEYLTGEKVRSPGEECDKVFTAIC 694

Qy        708 DGRI 711
               :|:|
Db        695 EGKI 698

RESULT 7
PAL4_ARATH
ID   PAL4_ARATH     STANDARD;      PRT;   707 AA.
AC   Q9SS45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   AT3G10340 OR F14P13.6.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AC009400; AAF02809.1; -.
DR   HSSP; P21310; 1B8F.
```

```
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    194    196       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     195    195       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   707 AA;  76919 MW;  B67743C941B50B9B CRC64;

  Query Match             28.3%;  Score 1020;  DB 1;  Length 707;
  Best Local Similarity   36.6%;  Pred. No. 1.4e-60;
  Matches  261;  Conservative  116;  Mismatches  248;  Indels   88;  Gaps   16;

Qy         49 SQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLK 108
              | |: |: :: :    : |:| | :||:  |     || |      |:  :| || |  | :::
Db         30 SHLDEVKRMVKEYRKEAVKLGGETLTIGQVAAVARGGGGSTVELAEEARAGVKASSEWVM 89

Qy        109 AQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRG-LENT 165
                 :      ||||||| ::  ||:   :||  ||      |:         | | |    :|
Db         90 ESMNRGTDSYGVTTGFGATSHRRTKQGGALQNELIRFLNAGI-------FGPGAGDTSHT 142

Qy        166 LPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLS 225
              ||      | ||:::|| :|:| :| ||:|:|   ||| |||  ||| :||||:|:||||| |||
Db        143 LPKPTTRAAMLVRVNTLLQGYSGIRFEILEAITKLLNHEITPCLPLRGTITASGDLVPLS 202

Qy        226 YIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEAVV-LGPKEGLGLVNGTAVSASM 284
              |||| :|| |:  |    |    | : : | ||  | |: :   | ||||| ||||||| : :
Db        203 YIAGLLTGRPNSKA-VGPSG--ETLTASEAFKLAGVSSFFELQPKEGLALVNGTAVGSGL 259

Qy        285 ATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSG 344
              |:  | |:::|::||: ::|:  | | |:    |    :     : ||||:| |   :  :| |
Db        260 ASTVLFDANILAVLSEVMSAMFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDG 318

Qy        345 SSFAVE----HEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTT 400
              ||: |      || :   |         :|||  ||||||:|||  :| :   |    :   |  |
Db        319 SSYVKEAQLLHEMDPLQKP-----KQDRYALRTSPQWLGPQIEVIRAATKMIEREINSVN 373

Qy        401 DNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSC 460
              ||||:||    :   ||||| : : : ::|: :||:|  ||||  | | :||:|    | ||||
Db        374 DNPLIDVSRNKALHGGNFQGTPIGVAMDNSRLAIASIGKLMFAQFSELVNDFYNNGLPSN 433

Qy        461 LA-AEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTA 519
              |:  :|||:|  || :|  :|:|  |||  ||||||  || ||  || ||||  ||:|:||
Db        434 LSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTA 493

Qy        520 EANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVK 579
              || |:|  |:   ::|      ||||| :|  :  ||                       ||
Db        494 EAVDILKLMSTTYLVALCQAVDLRHLEENLKK-------------------------AVK 528

Qy        580 KALNKRLEQTTTYDLEPRWHDA----------------FSYATG--------------TV 609
               |::: ::  |        | :                ||||                    :
Db        529 SAVSQVAKRVLTVGANGELHPSRFTERDVLQVVDREYVFSYADDPCSLTYPLMQKLRHIL 588

Qy        610 VELLSSSP--SANVTLTAVNAWKVASAEKAISLTREV-RNRFWQTPSSQAPAHAYLSPRT 666
              |:  :  |   ||  :   :       || :  | :|| | |         : | |:     |:
Db        589 VDHALADPEREANSATSVFHKIGAFEAELKLLLPKEVERVRVEYEEGTSAIANRIKECRS 648

Qy        667 RVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719
                 || |||:||        ::   |     : |      ::: || ||::    |:: |
Db        649 YPLYRFVRDELNT-----ELLTGENVRSPGEEFDKVFLAISDGKLIDPLLECL 696


RESULT 8
PALY_TRISU
ID   PALY_TRISU     STANDARD;      PRT;    725 AA.
AC   P45734;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL1.
OS   Trifolium subterraneum (Subterranean clover).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
OX   NCBI_TaxID=3900;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Karridale; TISSUE=Leaf;
RX   MEDLINE=94171061; PubMed=8125321;
RA   Howles P.A., Arioli T., Weinman J.J.;
RT   "Characterization of a phenylalanine ammonia-lyase multigene family
RT   in Trifolium subterraneum.";
RL   Gene 138:87-92(1994).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M91192; AAA17993.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    211    213       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     212    212       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   725 AA;  78998 MW;  70F9925C86622240 CRC64;

  Query Match             28.2%;  Score 1019;  DB 1;  Length 725;
  Best Local Similarity   36.0%;  Pred. No. 1.6e-60;
  Matches  273;  Conservative  127;  Mismatches  254;  Indels  104;  Gaps   22;

Qy          6 DSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDV 65
              ||   |  | ||   |         :|| |  :         |:     |  | |: |: :: :    |
Db         17 DSFCLTHAN--ANNMKV----NAADPLNW-----GVAAEAMKGSHLDEVKRMVEEYRKPV 65

Qy         66 VELSGYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGF 122
              | | | :||:  |    ||  |   |  :    :  ||  |  |  |::    :       ||||||
Db         66 VRLGGETLTISQVAAIAAHDGATVEL--SESARAGVKASSDWVMESMNKGTDSYGVTTGF 123

Qy        123 GGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSL 182
              | ::  ||:   :||| ||      |:         |   |      :|||      | ||::|:|:|
Db        124 GATSHRRTKQGGALQKELIRFLNAGI-------FGNGTESNHTLPHTATRAAMLVRINTL 176

Qy        183 TRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVL 242
               :|:| :|  :|||:|  ||:  |||  :||||:|:|||| |||||| |||||| :||   :  |
Db        177 LQGYSGIRFEILEAITKLLNNNITPCLPLRGTITASGDLVPLSYIAGLLTGRSNSKA--- 233

Qy        243 HEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              |  : ::: |:|| | |: |   |  ||||| ||||||| : :|:: |  ::::|::||:
Db        234 HGPSGEMLNAKEAFQLAGINAEFFELQPKEGLALVNGTAVGSGLASIVLFEANILAVLSE 293
```

```
Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVE----HEEEVK 356

Db        294 VLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILHGSAYVKDAKKLHEMDPL 352

Qy        357 VKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGG 416

Db        353 QKP-----KQDRYALRTSPQWLGPLIEVIRFSTKSIEREINSVNDNPLIDVSRNKALHGG 407

Qy        417 NFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGL 475

Db        408 NFQGTPIGVSMDNTRLALASIGKLLFAQFSELVNDFYNNGLPSNLSASRNPSLDYGFKGS 467

Qy        476 DIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYC 535

Db        468 EIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTKEAIEILQLMSSTFLIA 527

Qy        536 TLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLE 595

Db        528 LCQAIDLRHLEENLK--------------------NSVKNTVSQVAKKTLTIGVSGELH 566

Qy        596 P-------------RWHDAFSY------ATGTVVELL-----------SSSPSANVTLT 624

Db        567 PSRFCEKDLLKVVDREH-VFSYIDDPCSATYPLAQKLRQVLVDHALVNGESEKNSNTSIF 625

Qy        625 AVNAWKVASAEKAIS--LTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQA 681

Db        626 Q----KIATFEEELKTLLPKEVESARTAYENGNSTIANKINGCRSYPLYKFVREELGTSL 681

Qy        682 RRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719

Db        682 LTGERVI-----SPGEECDKLFTAMCQGKIIDPLLKCL 714

RESULT 9
PAL3_PETCR
ID   PAL3_PETCR     STANDARD;      PRT;   718 AA.
AC   P45729;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 3 (EC 4.3.1.5).
GN   PAL3.
OS   Petroselinum crispum (Parsley) (Petroselinum hortense).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;
OC   Apium clade; Petroselinum.
OX   NCBI_TaxID=4043;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95010141; PubMed=7925471;
RA   Appert C., Logemann E., Hahlbrock K., Schmid J., Amrhein N.;
RT   "Structural and catalytic properties of the four phenylalanine
RT   ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.).";
RL   Eur. J. Biochem. 225:491-499(1994).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X81159; CAA57057.1; -.
DR   PIR; S48726; S48726.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    204    206       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     205    205       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   718 AA;  78165 MW;  17786451FFDBDC35 CRC64;

  Query Match             28.2%;  Score 1017.5;  DB 1;  Length 718;
  Best Local Similarity   36.5%;  Pred. No. 2e-60;
  Matches  269;  Conservative  128;  Mismatches  275;  Indels   65;  Gaps   19;

Qy         14 NGFTNGSHAAPTKSAAGPTSALRRTP----GLDGHAAHQSQLEIVQELLSDPTDDVVELS 69

Db          5 NGTTNG-HA----NGNGLDLCMKKEDPLNWGVAAEALTGSHLDEVKRMVAEYRKPVVKLE 59

Qy         70 GYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSA 126

Db         60 GETLTISQVAAISARDDSGVKVELSEEARAGVKASSDWVMDSMNKGTDSYGVTTGFGATS 119

Qy        127 DTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE---NTLPLEVVRGAMVIRVNSLT 183

Db        120 HRRTKQGGALQKELIRFLNAGI---------FGSGAEAGNNTLPHSATRAAMLVRINTLL 170

Qy        184 RGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLH 243

Db        171 QGYSGIRFEILEAITKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA---V 227

Qy        244 EGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQA 301

Db        228 GPTGVTLSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEV 287

Qy        302 LTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDE 361

Db        288 MSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEMDPL 345

Qy        362 GILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQAS 421

Db        346 QKPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGS 405

Qy        422 AVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIA 480

Db        406 PIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMA 465

Qy        481 AYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAV 540

Db        466 SYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAI 525

Qy        541 DLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ----------- 588

Db        526 DLRHLEENLKSTVKNTVSQVAKRVLTMG--VNG-ELHPSRFCEKDLLRVVDREYIFAYID 582

Qy        589 ---TTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS--LTRE 643

Db        583 DPCSATYPLMQKLRE-------TLVEHALNNGDKERNLSTSIFQKIAAFEDELKALLPKE 635

Qy        644 VRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRI 702

Db        636 VETARAALESGNPAIPNRIKECRSYPLYKFVREELGT-----EYLTGEKVRSPGEEFEKV 690

Qy        703 YEAIKDGRINHVLVKML 719

Db        691 FTAMSKGEIIDPLLECL 707
```

```
RESULT 10
PAL1_TOBAC
ID   PAL1_TOBAC     STANDARD;      PRT;   715 AA.
AC   P25872;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   TPA1 OR PALB.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fukasawa-Akada T.;
RL   Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Bright Yellow; TISSUE=Callus;
RA   Taguchi G., Sharan M., Gonda K., Yanagisawa K., Shimosaka M.,
RA   Hayashida N., Okazaki M.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M84466; AAA34122.1; -.
DR   EMBL; AB008200; BAA22948.1; -.
DR   PIR; S66343; S66343.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    201    203       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     202    202       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   715 AA;  77780 MW;  03CB4E8527394C62 CRC64;

  Query Match            28.2%;  Score 1016.5;  DB 1;  Length 715;
  Best Local Similarity   36.0%;  Pred. No. 2.4e-60;
  Matches  270;  Conservative  130;  Mismatches  259;  Indels   91;  Gaps   19;

Qy         13 ANGFTN-GSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVELSGY 71
                 :||  | | :     | :| | :          | |      | |: |::::|:     :|:| |
Db          3 SNGHVNGGENFELCKKSADPLNWEMAAESLRG-----SHLDEVKKMVSEFRKPMVKLGGE 57

Qy         72 SLTVRDVVGAA---RKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSA 126
              ||||  |   |    :      |:|:  :| ||  |  | |::   :      |||||||| ::
Db         58 SLTVAQVAAIAVRDKSANGVKVELSEEARAGVKASSDWVMDSMNKGTDSYGVTTGFGATS 117

Qy        127 DTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGH 186
                 ||::  :|||  ||       ||          |  |     :|||    | ||::|:|:|  :|:
Db        118 HRRTKNGGALQKELIRFLNAGV-------FGNGTETSHTLPHSATRAAMLVRINTLLQGY 170

Qy        187 SAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGT 246
               | :|  :|||:|  :|   |||  :|||:|:||||| ||||||| :|| |:  |   |    |
Db        171 SGIRFEILEAITKLINSNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPNG- 228

Qy        247 EKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTA 304
               : : | ||  : |:         | ||||| ||||||| : ||:: | ||::|:::|: |:|
Db        229 -ETLNAEEAFRVAGVNGGFFELQPKEGLALVNGTAVGSGMASMVLFDSNILAVMSEVLSA 287

Qy        305 LTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGIL 364
               :  | | |:    |    :     : ||||:| |   :  :| |||: |:  :::    |
Db        288 IFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSY-VKAAQKLHEMDPLQKP 345

Qy        365 RQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVS 424
              :||||  ||||||:|||  :| :   |    :   |  |:  ||||:||     :  ||||| : :
Db        346 KQDRYALRTSPQWLGPQIEVIRAATKMIEREINSVNDNPLIDVSRNKALHGGNFQGTPIG 405

Qy        425 ISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGLDIHIAAYA 483
              :||:  |||||  ||||  |  |  :||:|      |||| |  |: :|||:| || :| :|:|
Db        406 VSMDNARLALASIGKLMFAQFSELVNDYYNNGLPSNLTASRNPSLDYGFKGAEIAMASYC 465

Qy        484 SELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLR 543
              |||  ||||||  || ||  || ||||  |||||:||||  |:| |:  :::|    ||:|||
Db        466 SELQFLANPVTNHVQSAEQHNQDVNSLGLISARKTAEAVDILKLMSSTYLVALCQAIDLR 525

Qy        544 AMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDA-- 601
               :| : |                    ||      ||  :::  ::| |          | |
Db        526 HLEENLK--------------------NA----VKNTVSQVAKRTLTMGANGELHPARF 560

Qy        602 -------------FSYATG-------------TVVELLSSSPSANVTLTAVNAWKVAS 633
                            |:||                 :|:   ::   :    :   |: :
Db        561 CEKELLRIVDREYLFAYADDPCSCNYPLMQKLRQVLVDHAMNNGESEKNVNSSIFQKIGA 620

Qy        634 AE---KAISLTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVG 689
                |   ||: | :|| : |      :  |  :     |:  || |||:|||         :: |
Db        621 FEDELKAV-LPKEVESARAALESGNPAIPNRITECRSYPLYRFVRKELGT-----ELLTG 674

Qy        690 VQQETIGSNVSRIYEAIKDGRINHVLVKML 719
               :  : |    :::  |: :|:|    ::: |
Db        675 EKVRSPGEECDKVFTAMCNGQIIDPMLECL 704

RESULT 11
PALY_POPTR
ID   PALY_POPTR     STANDARD;      PRT;   715 AA.
AC   P45730;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Populus trichocarpa (Western balsam poplar).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
OX   NCBI_TaxID=3694;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=P.trichocarpa X P.deltoides; TISSUE=Leaf;
RX   MEDLINE=94151434; PubMed=8108506;
RA   Subramaniam R., Reinold S., Molitor E.K., Douglas C.J.;
RT   "Structure, inheritance, and expression of hybrid poplar (Populus
RT   trichocarpa x Populus deltoides) phenylalanine ammonia-lyase genes.";
RL   Plant Physiol. 102:71-83(1993).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
```

```
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L11747; AAA33805.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    201    203       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     202    202       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   715 AA;  77918 MW;  5B9E837A6E8AF2B6 CRC64;

  Query Match             28.1%;  Score 1016;  DB 1;  Length 715;
  Best Local Similarity   35.8%;  Pred. No. 2.5e-60;
  Matches  266;  Conservative  132;  Mismatches  276;  Indels   68;  Gaps   18;

Qy          8 LATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVE 67
              : |   ||: |||  :    :   |  |        |:      | | |: |: :::|    ||:
Db          1 METVTKNGYQNGSLESLCVNQRDPLSW-----GVAAEAMKGSHLDEVKRMVADYRKPVVK 55

Qy         68 LSGYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGG 124
              | | :||:  |    |         |:|:  :  |   |   |  |::    :      |||||||
Db         56 LGGETLTIAQVASIAGHDTGDVKVELSESARPGVKASSDWVMDSMDKGTDSYGVTTGFGA 115

Qy        125 SADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE--NTLPLEVVRGAMVIRVNSL 182
              ::  ||:    :|||  ||      |:               | |  |   :|||    | ||::|:|:|
Db        116 TSHRRTKQGGALQKELIRFLNAGI---------FGNGTETCHTLPHSATRAAMLVRINTL 166

Qy        183 TRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVL 242
              :|:| :|  :|:||:|  ||: |||  :|||:|:||||  ||||||  :|| |: |
Db        167 LQGYSGIRFEILEAITRLLNNNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA--- 223

Qy        243 HEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQ 300
              | :::  |  ||    |:|:      | |||||  ||||||  : :|:: |  ::::|:|:||:
Db        224 TGPTGEVLDAAEAFKAAGIESGFFELQPKEGLALVNGTAVGSGLASMVLFETNVLAVLSE 283

Qy        301 ALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVE----HEEEVK 356
              |:|:  |  |  |:     |      :      : ||||:|  |    :  :|  ||::         || :
Db        284 LLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAYMKAAKKLHETDPL 342

Qy        357 VKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGG 416
              |          :||||  ||||||:|||  :|  :    ::       |  |:     ||||:||      :  |||
Db        343 QKP-----KQDRYALRTSPQWLGPQIEVIRFSTKSIEREINSVNDNPLIDVSRNKAIHGG 397

Qy        417 NFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGL 475
              |||  :  :  :||:  |||:|  ||||  |  |  :||:|    |  ||||  |  |:  :|||:|   ||
Db        398 NFQGTPIGVSMDNVRLAIASIGKLLFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKGA 457

Qy        476 DIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYC 535
              :|  :|:|  |||  :|||||||  ||  ||   ||  ||||  |||:|:||||  |:|  |:   :  |
Db        458 EIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSTTFLVA 517

Qy        536 TLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ------- 588
              ||:|||  :|  :          :  :  ::  |  ||  :          |        |  |  :
Db        518 LCQAIDLRHLEENLKSAVKNTVSQVSKRVLTTGANGE---LHPSRFCEKELLKVVDREYV 574

Qy        589 --------TTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS- 639
                     :  ||  |   :       |          |     |  |  :    :          |:|    |:  :  |:  :
Db        575 FAYVDDPCSATYPLMQKLRQVF------VDHALENGENEKNFSTSV-FQKIEAFEEELKA 627

Qy        640 -LTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGS 697
                   | :||  :  |          :  |    :          |:    ||  |||||||          :    |  :  ::  |
Db        628 LLPKEVESARAAYDSGNSAIDNKIKECRSYPLYKFVREELGTV-----LLTGEKVQSPGE 682

Qy        698 NVSRIYEAIKDGRINHVLVKML 719
                   :::  |:   |:|    :::  |
Db        683 EFDKVFTAMCQGKIIDPMLECL 704

RESULT 12
PAL2_ARATH
ID   PAL2_ARATH     STANDARD;      PRT;   717 AA.
AC   P45724; Q94KC9; Q9SCN5;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
GN   PAL2 OR AT3G53260 OR T4D2.190.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=95195160; PubMed=7888622;
RA   Wanner L.A., Li G., Ware D., Somssich I.E., Davis K.R.;
RT   "The phenylalanine ammonia-lyase gene family in Arabidopsis
RT   thaliana.";
RL   Plant Mol. Biol. 27:327-338(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=22954850; PubMed=14593172;
RA   Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA   Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
```

```
RA   Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA   Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA   Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA   Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA   Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA   Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA   Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA   Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA   Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Empirical analysis of transcriptional activity in the Arabidopsis
RT   genome.";
RL   Science 302:842-846(2003).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L33678; AAC18871.1; -.
DR   EMBL; AL132958; CAB64229.1; -.
DR   EMBL; AF367308; AAK32895.1; -.
DR   PIR; T46172; T46172.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    203    205       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     204    204       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
FT   CONFLICT    206    206       D -> H (IN REF. 1).
FT   CONFLICT    235    235       A -> E (IN REF. 1).
SQ   SEQUENCE   717 AA;  77859 MW;  360A206295AF787D CRC64;

  Query Match            28.1%;  Score 1015.5;  DB 1;  Length 717;
  Best Local Similarity  36.0%;  Pred. No. 2.8e-60;
  Matches  264;  Conservative  140;  Mismatches  284;  Indels   45;  Gaps   17;

Qy          5 LDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDD 64
Db          1 MDQIEAMLCGGGEKTKVAVTTKTLADPLNW-----GLAADQMKGSHLDEVKKMVEEYRRP 55

Qy         65 VVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGF 122
Db         56 VVNLGGETLTIGQVAAISTVGGSVKVELAETSRAGVKASSDWVMESMNKGTDSYGVTTGF 115

Qy        123 GGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSL 182
Db        116 GATSHRRTKNGTALQTELIRFLNAGI-------FGNTKETCHTLPQSATRAAMLVRVNTL 168

Qy        183 TRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVL 242
Db        169 LQGYSGIRFEILEAITSLLNHNISPSLPLRGTITASGDLVPLSYIAGLLTGRPNSKA--- 225

Qy        243 HEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
Db        226 -TGPDGESLTAKEAFEKAGISTGFFDLQPKEGLALVNGTAVGSGMASMVLFEANVQAVLA 284

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
Db        285 EVLSAIFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAIMEHILDGSSY-MKLAQKVHEMD 342

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQ 419
Db        343 PLQKPKQDRYALRTSPQWLGPQIEVIRQATKSIEREINSVNDNPLIDVSRNKAIHGGNFQ 402

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGLDIH 478
Db        403 GTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLTASSNPSLDYGFKGAEIA 462

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
Db        463 MASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLVGICQ 522

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRL------EQTTTY 592
Db        523 AVDLRHLEENLRQTVKNTVSQVAKKVLTTGINGE---LHPSRFCEKDLLKVVDREQVFTY 579

Qy        593 DLEP--RWHDAFSYATGTVVE-LLSSSPSANVTLTAVNAWKVASAEKAIS--LTREVRNR 647
Db        580 VDDPCSATYPLMQRLRQVIVDHALSNGETEKNAVTSI-FQKIGAFEEELKAVLPKEVEAA 638

Qy        648 FWQTPSSQAPAHAYLSP-RTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAI 706
Db        639 RAAYGNGTAPIPNRIKECRSYPLYRFVREELGTKLLTGEKVV-----SPGEEFDKVFTAM 693

Qy        707 KDGRINHVLVKML 719
Db        694 CEGKLIDPLMDCL 706


RESULT 13
PALY_DIGLA
ID   PALY_DIGLA     STANDARD;      PRT;   713 AA.
AC   O23924;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS   Digitalis lanata (Foxglove).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Lamiales; Antirrhinaceae; Digitalideae; Digitalis.
OX   NCBI_TaxID=49450;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Thoeringer C.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ002221; CAA05251.1; -.
```

```
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    198    200       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     199    199       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   713 AA;  77732 MW;  747CE64C65BDCC6A CRC64;

  Query Match             28.1%;  Score 1013.5;  DB 1;  Length 713;
  Best Local Similarity   36.1%;  Pred. No. 3.7e-60;
  Matches  256;  Conservative  122;  Mismatches  251;  Indels   81;  Gaps   14;

Qy         49 SQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLK 108
              | |: |: :: :      ||:| | :||:  |  :     | ||   :  || |   | |::
Db         35 SHLDEVKRMVEEFRKTVVKLGGETLTISQVAAIAARDNEVAVQLAESSRAGVKASSDWVM 94

Qy        109 AQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTL 166
                 :      |||||||| ::  ||:   :||| ||      |:         |  |    :||
Db         95 ESMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI-------FGNGTESTHTL 147

Qy        167 PLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSY 226
              |     | ||::|:|:| :|:| :|   :|| :| |||| ||| :||||:|:||||| ||||
Db        148 PHSATRAAMLVRINTLLQGYSGIRFEILETITKFLNHNITPCLPLRGTITASGDLVPLSY 207

Qy        227 IAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEAVV---LGPKEGLGLVNGTAVSAS 283
              ||| :|| |:  |    |      :  : |  :|  |  |  : :    ||||| |||||| :
Db        208 IAGLLTGRPNSKA-VGPNG--ESLNAEQAFKLAGANSGLFFELQPKEGLALVNGTAVGSG 264

Qy        284 MATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLS 343
              :|:::|::::|||:: ::|:  | | |:    |      :       : ||||:| |   :|
Db        265 LASIALYEANILSLLAEVMSAVFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILD 323

Qy        344 GSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNP 403
              |||: |:  :::   |        :||||  ||||||:|||  :|  :    |     :  | |:  |||
Db        324 GSSY-VKAAQKMHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRTATKMIEREINSVNDNP 382

Qy        404 LLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA- 462
              |:||   :   ||||| : : :||: :|||:| |||| | | :||:|   |  ||||  |:
Db        383 LIDVSRNKALHGGNFQGTPIGVSMDNSRLAIASIGKLMFAQFSELVNDFYNNGLPSNLSG 442

Qy        463 AEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEAN 522
              :|||:|  || :| :|:| |||  ||||||  ||   ||  || ||| |||:|:| ||
Db        443 GRNPSLDYGFKGSEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTVEAL 502

Qy        523 DVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKAL 582
              |:| |: :::|    ||:|||  :|                           | | || |   :
Db        503 DILKLMSSTYLVALCQAIDLRHLE-----------------------ENLRLSVKNTI 537

Qy        583 NKRLEQTTTYDLEPRWHDA----------------FSY------ATGTVVE--------- 611
              :: ::|  |  :       |  :                   |:|        ||  :::
Db        538 SQVAKRTLTTGVNGELHPSRFCELDLLRVVDREYVFAYVDDPCSATYPLMQKLRQVLVEH 597

Qy        612 -LLSSSPSANVTLTAVNAWKVASAEKAISLTREVRN-RFWQTPSSQAPAHAYLSPRTRVL 669
               |  :     | : :       ||   |  :|| :   |        | |:       |:  |
Db        598 ALKNGENEKNASTSIFQKIEAFEAELKAVLPKEVESARVALEDGKPAIANRITECRSYPL 657

Qy        670 YSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719
              | |:|||||        :      |  :  :  |      |:: |:  |  |  |:| |
Db        658 YKFIREELGT-----NFLTGEKVMSPGEECDRVFTAMSKGLIVDPLLKCL 702

RESULT 14
PALY_CAMSI
ID   PALY_CAMSI     STANDARD;      PRT;   714 AA.
AC   P45726;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Camellia sinensis (Tea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   Ericales; Theaceae; Camellia.
OX   NCBI_TaxID=4442;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Yabukita; TISSUE=Leaf;
RA   Matsumoto S., Takeuchi A., Hayatsu M., Kondo S.;
RT   "Molecular cloning of phenylalanine ammonia-lyase cDNA and
RT   classification of varieties and cultivars of tea plants (Camellia
RT   sinensis) using the tea PAL cDNA probe.";
RL   Theor. Appl. Genet. 89:671-675(1994).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D26596; BAA05643.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    200    202       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     201    201       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   714 AA;  77751 MW;  5F0EC7E8D0AB9071 CRC64;

  Query Match             28.0%;  Score 1012;  DB 1;  Length 714;
  Best Local Similarity   36.2%;  Pred. No. 4.7e-60;
  Matches  268;  Conservative  130;  Mismatches  280;  Indels   62;  Gaps   18;

Qy          5 LDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDD 64
              :||  | : ||   || :|      | :         |:   |    | || |: :: :
Db          1 MDS-TTAIGNGV--GSGGSPGFCLKDPLNW-----GVAAEAMKGSHLEEVKGMVEEFRKP 52

Qy         65 VVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGF 122
              || | :||:  |   |:  :|   |  |: |   |    |:  :        :     ||||||
Db         53 VVRLGGETLTISQVAAIAVRGSEVAVELSESAREGVKASSDWVMESMNKGTDSYGVTTGF 112

Qy        123 GGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE--NTLPLEVVRGAMVIRVN 180
              | ::  ||::  :|| ||          |:             | |  |   :|||     | ||::|:|
Db        113 GATSHRRTKEGGALQKELIRFLNAGI---------FGNGTESCHTLPQSATRAAMLVRIN 163

Qy        181 SLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVH 240
              :| :|:| :|  :  :||:: |||: ||| :||||:|:||||| |||||| :|| :   |
Db        164 TLLQGYSGIRFEILEAISKFLNNNITPCLPLRGTITASGDLVPLSYIAGLLTGRHNSKA- 222

Qy        241 VLHEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLL 298
                   | :|:  :||  | |:|       | ||||| ||||||| : :|:: | ::::|::|
Db        223 --VGPTGEILHPKEAFRLAGVEGGFFELQPKEGLALVNGTAVGSGLASMVLFEANILAVL 280
```

```
Qy        299 SQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVK 358
Db        281 SEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSY-VKAAQKLHEM 338

Qy        359 DDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNF 418
Db        339 DPLQKPKQDRYALRTSPQWLGPLIEVIRSSTKSIEREINSVNDNPLINVSRNKALHGGNF 398

Qy        419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDI 477
Db        399 QGTPIGVSMDNTRLAVASIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEI 458

Qy        478 HIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTL 537
Db        459 AMAAYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSSTYLVALC 518

Qy        538 QAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ--------- 588
Db        519 QAVDLRHFEENLRNTVKSTVSQVAKRVLTMG--VNG-ELHPSRFCEKDLLRVVDREYIFA 575

Qy        589 ------TTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS--L 640
Db        576 YIDDPCSATYPLMQKLRQ-------VLVEHALKNGESEKNLSTSIFQKIRAFEEEIKTLL 628

Qy        641 TREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNV 699
Db        629 PKEVESTRAAIENGNSAIPNRIKECRSYPLYKFVREELGT-----ELLTGEKVRSPGEEF 683

Qy        700 SRIYEAIKDGRINHVLVKML 719
Db        684 DKVFTALCKGEMIDPLMDCL 703
```

```
RESULT 15
PAL1_SOYBN
ID   PAL1_SOYBN     STANDARD;      PRT;   713 AA.
AC   P27991;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN   PAL1.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. T225; TISSUE=Leaf;
RX   MEDLINE=92190550; PubMed=1799682;
RA   Frank R.L., Vodkin L.O.;
RT   "Sequence and structure of a phenylalanine ammonia-lyase gene from
RT   Glycine max.";
RL   DNA Seq. 1:335-346(1991).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X52953; CAA37129.1; -.
DR   PIR; S22991; S22991.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    199    201       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     200    200       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   713 AA;  77744 MW;  9D71EF1CC230216A CRC64;

  Query Match             28.0%;  Score 1011;  DB 1;  Length 713;
  Best Local Similarity   36.9%;  Pred. No. 5.5e-60;
  Matches  270;  Conservative  120;  Mismatches  282;  Indels   60;  Gaps   19;

Qy         14 NGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVELSGYSL 73
Db          5 NGHQNGSFCLST--AKGNNDPLNW--GAAAEAMKGSHLDEVKRMVAEYRKPVVRLGGETL 60

Qy         74 TVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRTE 131
Db         61 TIAQVAAVAGHDHGVAVELSESAREGVKASSEWVMNSMNNGTDSYGVTTGFGATSHRRTK 120

Qy        132 DAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVRL 191
Db        121 QGGALQKELIRFLNAGI-------FGNGTESSHTLPHTATRAAMLVRINTLLQGYSGIRF 173

Qy        192 VVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMF 251
Db        174 EILEAITKLLNNNVTPCLDLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPSG--EVLN 230

Qy        252 AREAISLFGL--EAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALTVEA 309
Db        231 AKEAFELASINSEFFELQPKEGLALVNGTAVGSGLASMVLFEANILAVLSEVLSAIFAEV 290

Qy        310 MVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVE----HEEEVKVKDDEGILR 365
Db        291 MQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSYMKAAKKLHEIDPLQKP-----K 344

Qy        366 QDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVSI 425
Db        345 QDRYALRTSPQWLGPLIEVIRFSTKSIEREINSVNDNPLIDVSRNKALHGGNFQGTPIGV 404

Qy        426 SMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGLDIHIAAYAS 484
Db        405 SMDNTRLALASIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCS 464

Qy        485 ELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLRA 544
Db        465 ELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTNEAIEILKLMSSTFLIALCQAIDLRH 524

Qy        545 MELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDAFSY 604
Db        525 LEENLKNSVKNTVSQVSKRILTTG--VNG-ELHPSRFCEKDLLKVVDREY------IFSY 575

Qy        605 ---ATGTVVELLSSSPSANVTLTAVNA----------WKVASAEKAIS--LTREVRNRF 648
Db        576 IDDPCSATYPLMQKLRQVLVDHALVNAECEKDVNSSIFQKIAIFEEELKNLLPKEVEGAR 635

Qy        649 WQTPSSQAPAHAYLSP-RTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIK 707
Db        636 AAYESGKAAIPNKIQECRSYPLYKFVREELGT-----GLLTGEKVRSPGEEFDKLFTAMC 690

Qy        708 DGRINHVLVKML 719
Db        691 QGKIIDPLMECL 702
```

Search completed: September 9, 2004, 10:04:39
Job time : 19.4315 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:57:28 ; Search time 87.6349 Seconds
(without alignments)
2592.269 Million cell updates/sec

Title: US-09-939-408A-13
Perfect score: 3610
Sequence: 1 MAPSLDSLATTLANGFTNGS..........RIYEAIKDGRINHVLVKMLA 720

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1041.5	28.9	725	10	Q94AN1	Q94an1 arabidopsis
2	1031.5	28.6	730	10	Q9M567	Q9m567 rubus idaeu
3	1025	28.4	711	10	Q8W2E4	Q8w2e4 lactuca sat
4	1018	28.2	711	10	Q9AXI5	Q9axi5 pharbitis n
5	1016.5	28.2	707	10	Q7X7G3	Q7x7g3 oryza sativ
6	1016.5	28.2	717	10	Q8RWP4	Q8rwp4 arabidopsis
7	1014.5	28.1	708	10	Q94EN0	Q94en0 rehmannia g
8	1013	28.1	715	10	Q852S1	Q852s1 daucus caro
9	1010.5	28.0	714	10	Q8H6V6	Q8h6v6 populus tre
10	1007.5	27.9	714	10	Q7X6M8	Q7x6m8 oryza sativ
11	1003.5	27.8	671	10	Q84VE0	Q84ve0 oryza sativ
12	1002.5	27.8	721	10	Q9XFX5	Q9xfx5 citrus clem
13	998	27.6	713	10	Q8GZR8	Q8gzr8 lactuca sat
14	997.5	27.6	703	10	Q8VXG7	Q8vxg7 zea mays (m
15	997	27.6	716	10	Q9ATN7	Q9atn7 agastache r
16	994	27.5	717	10	Q8H6W0	Q8h6w0 coffea cane
17	991.5	27.5	716	10	Q9MAX1	Q9max1 catharanthu
18	990.5	27.4	710	10	Q94C45	Q94c45 manihot esc
19	990	27.4	715	10	O24266	O24266 populus kit
20	984.5	27.3	687	10	Q94F90	Q94f90 manihot esc
21	982	27.2	711	10	Q8H6V5	Q8h6v5 populus tre
22	980.5	27.2	712	10	Q94F89	Q94f89 manihot esc
23	971.5	26.9	655	10	Q9XGR3	Q9xgr3 vigna ungui
24	967.5	26.8	718	10	Q9XFX6	Q9xfx6 citrus clem
25	957.5	26.5	703	10	Q84LI2	Q84li2 phalaenopsi
26	957	26.5	681	10	Q8RUZ3	Q8ruz3 pinus sylve
27	956	26.5	681	10	Q8SAS6	Q8sas6 pinus sylve
28	956	26.5	681	10	Q8RV49	Q8rv49 pinus sylve
29	955	26.5	549	10	O04876	O04876 hordeum vul
30	950	26.3	681	10	Q8SAS7	Q8sas7 pinus sylve
31	946	26.2	619	10	Q39545	Q39545 cucumis mel
32	936	25.9	619	10	Q8H6V9	Q8h6v9 coffea cane
33	934	25.9	618	10	Q9FS00	Q9fs00 dianthus ca
34	930.5	25.8	727	10	Q7XJC3	Q7xjc3 pinus pinas
35	929.5	25.7	727	10	Q7XJC4	Q7xjc4 pinus pinas
36	786	21.8	396	10	Q8VWP6	Q8vwp6 beta vulgar
37	780.5	21.6	532	10	O81647	O81647 capsicum ch
38	769.5	21.3	544	10	Q9FT38	Q9ft38 triticum ae
39	741.5	20.5	362	10	Q9FUI7	Q9fui7 rubus idaeu
40	741.5	20.5	380	10	Q84UZ7	Q84uz7 ginkgo bilo
41	714.5	19.8	363	10	Q9FUI8	Q9fui8 rubus idaeu
42	659.5	18.3	497	10	O04869	O04869 hordeum vul
43	655.5	18.2	475	10	O04875	O04875 hordeum vul
44	620	17.2	435	10	Q9M3U1	Q9m3u1 betula verr
45	619	17.1	436	10	Q9LEP1	Q9lep1 betula verr

ALIGNMENTS

RESULT 1
Q94AN1
ID Q94AN1 PRELIMINARY; PRT; 725 AA.
AC Q94AN1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia lyase (Hypothetical protein) (EC 4.3.1.5).
GN AT2G37040 OR PAL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene T1J8.22/At2g37040 (GI:6598547).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,
RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,
RA Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Shinn P., Tang C.C., Toroumi M., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,
RA Davis R.W.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,

```
RA   Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
RA   Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
RA   Southwick A., Davis R.W., Ecker J.R., Theologis A.;
RT   "Arabidopsis Open Reading Frame (ORF) Clones.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Cochrane F.C., Kim M.R., Cardenas C.L., Davin L.B., Lewis N.G.;
RT   "Putative PAL multigene family in Arabidopsis thaliana: Kinetic
RT   Characterization of Recombinant Proteins and Knock-out Mutant
RT   Characterization.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY045919; AAK76593.1; -.
DR   EMBL; BT003330; AAO29949.1; -.
DR   EMBL; AY079363; AAL85094.1; -.
DR   EMBL; AY303128; AAP59438.1; -.
DR   PIR; G84787; G84787.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Hypothetical protein; Lyase.
SQ   SEQUENCE   725 AA;  78725 MW;  02626B3B2DEFE9CE CRC64;

  Query Match             28.9%;  Score 1041.5;  DB 10;  Length 725;
  Best Local Similarity   36.5%;  Pred. No. 3.2e-61;
  Matches  272;  Conservative  138;  Mismatches  258;  Indels   77;  Gaps   21;

Qy         18 NGSHAAPTKSAAGPTSAL-------------RRTPGLDGHAAHQ---SQLEIVQELLSD 60
Db          4 NGAH----KSNGGGVDAMLCGGDIKTKNMVINAEDPLNWGAAAEQMKGSHLDEVKRMVAE 59

Qy         61 PTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGV 118
Db         60 FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARAGVNASSDWVMESMNKGTDSYGV 119

Qy        119 TTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIR 178
Db        120 TTGFGATSHRRTKNGVALQKELIRFLNAGI-------FGSTKETSHTLPHSATRAAMLVR 172

Qy        179 VNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVK 238
Db        173 INTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTITASGDLVPLSYIAGLLTGRPNSK 232

Qy        239 VHVLHEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHML 295
Db        233 A----TGPNGEALTAEEAFKLAGISSGFFDLQPKEGLALVNGTAVGSGMASMVLFETNVL 288

Qy        296 SLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEV 355
Db        289 SVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAIMEHILDGSSY-MKLAQKL 346

Qy        356 KVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHG 415
Db        347 HEMDPLQKPKQDRYALRTSPQWLGPQIEVIRYATKSIEREINSVNDNPLIDVSRNKAIHG 406

Qy        416 GNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKG 474
Db        407 GNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKG 466

Qy        475 LDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLY 534
Db        467 AEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLV 526

Qy        535 CTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRL------EQ 588
Db        527 AICQAVDLRHLEENLRQTVKNTVSQVAKKVLTTG--VNG-ELHPSRFCEKDLLKVVDREQ 583

Qy        589 TTTYDLEPRWHDAFSYATGTVVE----------LLSSSPSANVTLTAVNAWKVASAEKAI 638
Db        584 VYTYADDP------CSATYPLIQKLRQVIVDHALINGESEKNAVTSIFH--KIGAFEEEL 635

Qy        639 S--LTREVR--NRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQET 694
Db        636 KAVLPKEVEAARAAYDNGTSAIP-NRIKECRSYPLYRFVREELGT-----ELLTGEKVTS 689

Qy        695 IGSNVSRIYEAIKDGRINHVLVKML 719
Db        690 PGEEFDKVFTAICEGKIIDPMMECL 714


RESULT 2
Q9M567
ID   Q9M567      PRELIMINARY;      PRT;   730 AA.
AC   Q9M567;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
GN   PAL2.
OS   Rubus idaeus (Raspberry).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Rosales; Rosaceae; Rosoideae; Rubus.
OX   NCBI_TaxID=32247;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21437948; PubMed=11553751;
RA   Kumar A., Ellis B.E.;
RT   "The phenylalanine ammonia-lyase gene family in raspberry. structure,
RT   expression, and evolution.";
RL   Plant Physiol. 127:230-239(2001).
DR   EMBL; AF237955; AAF40224.1; -.
DR   HSSP; P21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   730 AA;  79422 MW;  0E942417E50CE113 CRC64;

  Query Match             28.6%;  Score 1031.5;  DB 10;  Length 730;
  Best Local Similarity   37.1%;  Pred. No. 1.5e-60;
  Matches  268;  Conservative  130;  Mismatches  276;  Indels   49;  Gaps   17;

Qy         14 NGFTNGS--HAAPTKSAAGPTSALRRTPGLDGHAAHQ---SQLEIVQELLSDPTDDVVEL 68
Db         13 NGIQNGSLDDGLCIKTESIKTGYSVSDPLNWGAAAESMTGSHLDEVRRMVAEYRKPVVKL 72

Qy         69 SGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSA 126
Db         73 GGETLTISQVAAIANHDSGVKVELAESARAGVKASSDWVMDSMNKGTDSYGVTTGFGATS 132

Qy        127 DTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGH 186
Db        133 HRRTKQGAALQKELIRFLNAGVLRNGTES-------AHTLPHSATRAAMLVRINTLLQGY 185

Qy        187 SAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGT 246
Db        186 SGIRFEILEAISKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPKG- 243
```

```
Qy        247 EKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTA 304
              : : | || :  |: :      | |||| ||||||| : :|:  | :::|:|||: |:|
Db        244 -ETLNAAEAFAQVGISSGFFELQPKEGLALVNGTAVGSGLASTVLFETNILALLSEILSA 302

Qy        305 LTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGIL 364
              : | | |:  |   :   : ||||:| |  :  :| |||: |: |:: :|
Db        303 IFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSY-VKAAEKLHEQDPLQKP 360

Qy        365 RQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVS 424
              :|||| ||||||:||| :| :  :  ::  | |:  ||||:||  :  ||||| : :
Db        361 KQDRYALRTSPQWLGPQIEVIRFSTKSIEREINSVNDNPLIDVSRNKALHGGNFQGTPIG 420

Qy        425 ISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAAYA 483
              :||: ||||:| |||| | :|:|  | |||| |:  ||||:|  ||: || :|:|
Db        421 VSMDNTRLAIASIGKLMFAQFSELVNDFYNNGLPSNLSGGRDPSLDYGFKGAEIAMASYC 480

Qy        484 SELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLR 543
              ||| ||||||  || ||  || |||| |||:|:||||:|| |: ::|  ||:|||
Db        481 SELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSSTFLVALCQAIDLR 540

Qy        544 AMELDFKKQFDPLLPTLLQQHLGTGL----------DVNALALEVKKALNKRLEQ--TTT 591
              :| : |    : | ::| ||:      : : | : ::|  :: :|
Db        541 HLEENLKSTVKNTVSQLAKRVLTTGVNGELHPSRFCEKDLLMVVEREYLFAYIDDPCSAT 600

Qy        592 YDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS--LTREVRN-RF 648
              | |  |  |  | ||:: :  |:: |:|:|: |:|| | :||:|
Db        601 YPLMQRLRQVL------VEHALTNGENEKNASTSI-FQKITAFEEELKTILPKEVESARA 653

Qy        649 WQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKD 708
              : | : : : |: || ||||||| |:  | : | :  | : | :::|:|
Db        654 AYESGNAAIPNRIVECRSYPLYKFVREELG-----GEFLTGEKVRSPGEECDKVFTAMCQ 708

Qy        709 GRI 711
              | |
Db        709 GNI 711
```

RESULT 3
Q8W2E4

```
ID   Q8W2E4      PRELIMINARY;      PRT;    711 AA.
AC   Q8W2E4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
OS   Lactuca sativa (Garden lettuce).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;
OC   Lactuca.
OX   NCBI_TaxID=4236;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Campos-Vargas R., Nonogaki H., Suslow T., Saltveit M.;
RT   "Characterization of phenylalanine ammonia-lyase (PAL) gene in wounded
RT   lettuce leaf tissue.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF299330; AAL55242.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   711 AA;  77363 MW;  3120AB2877C29513 CRC64;
```

```
  Query Match             28.4%;  Score 1025;  DB 10;  Length 711;
  Best Local Similarity   37.9%;  Pred. No. 4e-60;
  Matches  267;  Conservative  127;  Mismatches  259;  Indels   52;  Gaps   18;

Qy         40 GLDGHAAHQSQLEIVQELLSDPTDDVVELSGYSLTVRDVVG--AARKGRRVRVQNDDEIR 97
              |:   |    | |: |::::::     ||:| | :|||  |  |   ||      |:|:  :  |
Db         23 GVAAEALTGSHLDEVKKMVAEFRKPVVKLGGETLTVSQVAGIAAANDSDTVKVELSEAAR 82

Qy         98 ARVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSS 155
              |  |  |  |::  :        ||||||| ::  ||:   :||| ||      |:
Db         83 AGVKASSDWVMESMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI------- 135

Qy        156 FSVGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSI 215
              |  |    :|||    | ||::|:|:|:| :|:| :|  :|||:| |||: ||| :||||:|
Db        136 FGNGTETSHTLPHSATRAAMIVRINTLLQGYSGIRFEILEAITKFLNNNITPCLPLRGTI 195

Qy        216 SASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEA--VVLGPKEGLG 273
              :||||| |||||||  :|| |: |          | :::  | :| :  |:|   | |||||
Db        196 TASGDLVPLSYIAGLLTGRPNSKA---VGPTGEVLNAEKAFAAAGVEGGFFELQPKEGLA 252

Qy        274 LVNGTAVSASMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVE 333
              |||||| : ||:: | |:::|:|||: |:|:  |  | |:  |   :   : ||||:|
Db        253 LVNGTAVGSGMASMVLFDANVLALLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIE 311

Qy        334 VARNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLS 393
              |  :  :||  : : |:  | :  : | :  ::|  :|||| |||||:||| :| :  :  :
Db        312 AAAIMEYILDGSDY-VKAAQKVHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRSSTKMIE 370

Qy        394 LENNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAM 453
              | |: ||||:||  :  |||||| : : :||: ||||:| |||| | | :||:|
Db        371 REINSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFY 430

Qy        454 NRGLPSCLA-AEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLAL 512
              | |||| |:  :|||:| || :| :|:| ||| ||||||  || ||  || |||| |
Db        431 NNGLPSNLSGGRNPSLDYGFKGGEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGL 490

Qy        513 ISARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVN 572
              |||:|||| |:| |: :::|  :||||| |   :  : ::  |  |    |  ||
Db        491 ISARKTAEAVDILKLMSSTYLVALCQSIDLRHLEENMKSTVKNTVSQVAKKVLTMG--VN 548

Qy        573 ALALEVKKALNKRLEQTTTYDLEPRWHDAFSY----ATGT-----------VVELLSSSP 617
              |  :  | |  :    |:|    :||                |  |::
Db        549 G-ELHPSRFCEKDLLRVVDREY------VFAYIDDVCSGTYPLMQKLRQVLVDHALNNGE 601

Qy        618 SANVTLTAVNAWKVASAEKAIS--LTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVR 674
              :  | |::  |:|: |: :  | :||  |   : :  :  :  |:  || |||
Db        602 TEKNTNTSI-FQKIATFEEELKVLLPKEVEGVRIAYENDTLSIPNRIKACRSYPLYRFVR 660

Qy        675 EELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719
              ||||    ||    | :  :  |    |:: |:  |:|   |:: |
Db        661 EELG----RG-FLTGEKVTSPGEEFDRVFTAMCKGQIIDPLLECL 700
```

RESULT 4
Q9AXI5

```
ID   Q9AXI5      PRELIMINARY;      PRT;    711 AA.
AC   Q9AXI5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS   Pharbitis nil (Violet) (Japanese morning glory).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Convolvulaceae; Ipomoea.
OX   NCBI_TaxID=35883;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Violet;
RA   Nakazawa A., Nozue M., Yasuda H., Takeba G., Kubo H.;
RT   "Expression of phenylalanine ammonia-lyase in Pharbitis nil.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AF325496; AAG49585.1; -.
DR   HSSP; P21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   711 AA;  77085 MW;  CBE6E3A0C0969E90 CRC64;

  Query Match             28.2%;  Score 1018;  DB 10;  Length 711;
  Best Local Similarity   37.9%;  Pred. No. 1.2e-59;
  Matches  270;  Conservative  114;  Mismatches  241;  Indels   88;  Gaps   18;

Qy         49 SQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLK 108
Db         34 SHLDEVKVMVAEFRKPAVKLGGETLTVAQVAAIASRDNAVTVELSEESRAGVKASSDWVM 93

Qy        109 AQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE--N 164
Db         94 DSMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI---------FGNGTESCH 144

Qy        165 TLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPL 224
Db        145 TLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKLLNHNITPCLPLRGTITASGDLVPL 204

Qy        225 SYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSA 282
Db        205 SYIAGLITGRPNSKA-VGPNG--ETLNAEEALRLAGVNGGFFELQPKEGLALVNGTAVGS 261

Qy        283 SMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLL 342
Db        262 GMASMVLFEANVLAVLSEVLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHIL 320

Qy        343 SGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDN 402
Db        321 DGSSY-VKAAQKMHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRAATKMIEREINSVNDN 379

Qy        403 PLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL- 461
Db        380 PLIDVARSKALHGGNFQGTPIGVSMDNSRLALASIGKLLFAQFSELVNDYYNNGLPSNLT 439

Qy        462 AAEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEA 521
Db        440 AGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISARKTAEA 499

Qy        522 NDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKA 581
Db        500 VDVLKLMSSTYLVALCQAIDLRFLEENLR---------------------NA----VKNA 534

Qy        582 LNKRLEQTTTYDLEPRWHDA----------------FSYA---------------TGTVV 610
Db        535 VTQVAKRTLTMGANGELHPARFCEKDLLRVVDREYVFAYADDPCSANYPLMQKLRQALVD 594

Qy        611 ELLSSSPSANVTLTAVNAWKVASAE---KAISLTREVR-NRFWQTPSSQAPAHAYLSPRT 666
Db        595 HALQNGESEKNTGTSI-FLKVAAFEDELKAV-LPKEVEARIAVESGNPAIPNRIKECRS 652

Qy        667 RVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719
Db        653 YPLYKFVREGLGT-----ELLTGEKVRSPGEECDKVFTAMCEGSIIDPLLECL 700

RESULT 5
Q7X7G3
AC   Q7X7G3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   OSJNBa0073E02.12 protein (OSJNBb0061C13.1 protein).
GN   OSJNBA0073E02.12 OR OSJNBB0061C13.1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA   Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA   Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA   Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA   Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA   Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA   Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA   Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA   Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL731616; CAE05452.1; -.
DR   EMBL; AL731629; CAE05619.1; -.
SQ   SEQUENCE   707 AA;  76038 MW;  841E4C46D077B1AA CRC64;

  Query Match             28.2%;  Score 1016.5;  DB 10;  Length 707;
  Best Local Similarity   36.9%;  Pred. No. 1.5e-59;
  Matches  268;  Conservative  144;  Mismatches  254;  Indels   61;  Gaps   20;

Qy         22 AAPTKSAAGPTSALRRTPGLDGHAAHQ---SQLEIVQELLSDPTDDVVELSGYSLTVRDV 78
Db          2 ASQTADATGFVAS---DPLSWGKAALEMTGSHLDEVKRMVAQSREAVVKIEGSSLRVGQV 58

Qy         79 --VGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQN--SVYGVTTGFGGSADTRTEDAV 134
Db         59 AAVSAAKDASGVVVELDEEARPRVKASSEWILNCIAHGGDIYGVTTGFGGTSHRRTKDGQ 118

Qy        135 SLQKALIEHQLCGVTPTSXSSFSVGRGLE-NTLPLEVVRGAMVIRVNSLTRGHSAVRLVV 193
Db        119 ALQVELLRHLNAGI---------FGNGSDGNSLPSEVSRAAMLVRINTLLQGYSGIRFEI 169

Qy        194 LEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFAR 253
Db        170 LEAITKLINTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQA-VTVDG--KKVDAA 226

Qy        254 EAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALTVEAMV 311
Db        227 EAFKIAGIQGGFFRLEPKEGLAIVNGTSVGSALAAMVLYDANVLAVLSEVLSAVFCEVMN 286

Qy        312 GQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDRYPL 371
Db        287 GKP-EYTDHLTHKLKHHPGSIEAAAIMEHILAGSAF-MPHAQKVNEVDPLLKPKQDRYAL 344

Qy        372 RTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTR 431
Db        345 RTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNTR 404

Qy        432 LALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAAYASELGHLA 490
Db        405 LAIANIGKLMFAQFSELVNEFYNNGLTSNLAGSRNPSLDYGFKGTEIAMASYCSELQFLA 464

Qy        491 NPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFK 550
Db        465 NPVTNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYLVALCQAVDLRHLEENLK 524

Qy        551 KQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDAFSYATG--- 607
Db        525 SAVKNCVTTVAKKVLTTG---PAGGLHSARFSEKAL--LTAIDREA----VYSYADDPCS 575

Qy        608 ------------TVVELLSSSPSANVTLTAVNAWKVASAEKAI--SLTREVRNRFWQTPS 653
```

```
Db        576 ANYPLMTKIRAVLVEHALANGPAEKDDGSSVFS-KITAFEEELREALPREMEAARVAFET 634

Qy        654 SQAP-AHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRIN 712
                || :        |:  || ||||||         | : :: |     :::: || : ::
Db        635 GTAPITNRIKESRSFPLYRFVREELGCV-----YLTGEKLKSPGEECNKVFLAISERKLI 689

Qy        713 HVLVKML 719
                ::: |
Db        690 DPMLECL 696

RESULT 6
Q8RWP4
ID   Q8RWP4      PRELIMINARY;      PRT;   717 AA.
AC   Q8RWP4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
GN   AT3G53260.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY092957; AAM12956.1; -.
DR   EMBL; BT000035; AAN15354.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   717 AA;  77801 MW;  C4C9CBDCDB91A838 CRC64;

  Query Match            28.2%;  Score 1016.5;  DB 10;  Length 717;
  Best Local Similarity  36.0%;  Pred. No. 1.5e-59;
  Matches  264;  Conservative  141;  Mismatches  283;  Indels   45;  Gaps   17;

Qy          5 LDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDD 64
              :| :    |   |          | ||: | | :        ||         | |: |:::::
Db          1 MDQIEAMLCGGGEKTKVAVTTKTLADPLNW-----GLAADQMKGSHLDEVKKMVAEYRRP 55

Qy         65 VVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGF 122
              || | | :||:  |    :  |  |:|:. :  ||  |  | |::   :      ||||||
Db         56 VVNLGGETLTIGQVAAISTVGGSVKVELAETSRAGVKASSDWVMESMNKGTDSYGVTTGF 115

Qy        123 GGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSL 182
              | ::  ||::  :|| ||      |:       |   :  :|||     | ||::|||:|
Db        116 GATSHRRTKNGTALQTELIRFLNAGI-------FGNTKETCHTLPQSATRAAMLVRVNTL 168

Qy        183 TRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVL 242
              :|:| :|  :|||:|: ||| |:| :|||:|:||||| |||||| :|| |: |
Db        169 LQGYSGIRFEILEAITSLLNHNISPSLPLRGTITASGDLVPLSYIAGLLTGRPNSKA--- 225

Qy        243 HEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299
                | : : : |:||    |:       | ||||| |||||| : ||:: | :::: ::|:
Db        226 -TGPDGESLTAKEAFEKAGISTGFFDLQPKEGLALVNGTAVGSGMASMVLFEANVQAVLA 284

Qy        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359
              : |:|:  | | |:     |    :  |||:| |   :  :| ||: ::  ::|   |
Db        285 EVLSAIFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAIMEHILDGSSY-MKLAQKVHEMD 342

Qy        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQ 419
                 :|||| ||||||:|||  :| :   |   ::   | |:  ||||:||    :  |||||
Db        343 PLQKPKQDRYALRTSPQWLGPQIEVIRQATKSIEREINSVNDNPLIDVSRNKAIHGGNFQ 402

Qy        420 ASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYHGKGLDIH 478
              : : :||: ||||:| |||| | | :||:|   | ||||  | |: :|||:|  || :|
Db        403 GTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLTASSNPSLDYGFKGAEIA 462

Qy        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538
              :|:| ||| :||||||: || ||  ||  ||||  ||| |||:|:|:| |:| |:  :  |
Db        463 MASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLVGICQ 522

Qy        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRL------EQTTTY 592
              ||||| :| : ::           :  : :: | ||::       |    :      | |       || ||
Db        523 AVDLRHLEENLRQTVKNTVSQVAKKVLTTGINGE---LHPSRFCEKDLLKVVDREQVFTY 579

Qy        593 DLEP--RWHDAFSYATGTVVE-LLSSSPSANVTLTAVNAWKVASAEKAIS--LTREVRNR 647
                :|     :            :|:  ||:   :    :|::   |: : |: :   | :||
Db        580 VDDPCSATYPLMQRLRQVIVDHALSNGETEKNAVTSI-FQKIGAFEEELKAVLPKEVEAA 638

Qy        648 FWQTPSSQAPAHAYLSP-RTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAI 706
                 :  ||      :      |:  || |||||||  :     |:  |        :  |     ::: |:
Db        639 RAAYGNGTAPIPNRIKECRSYPLYRFVREELGTKLLTGEKVV-----SPGEEFDKVFTAM 693

Qy        707 KDGRINHVLVKML 719
                :|::   |:  |
Db        694 CEGKLIDPLMDCL 706

RESULT 7
Q94EN0
ID   Q94EN0      PRELIMINARY;      PRT;   708 AA.
AC   Q94EN0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL1.
OS   Rehmannia glutinosa.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Lamiales; Lamiales incertae sedis; Rehmannia.
OX   NCBI_TaxID=99300;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Keumsan;
RA   Yun S.J., Lee B.K., Park M.R., Srinivas B., Chun J.-C.;
RT   "Molecular characterization of phenylalanine ammonia-lyase gene in
RT   Rhemannia glutinosa.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF401636; AAK84225.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
```

```
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   Lyase.
SQ   SEQUENCE   708 AA;  76835 MW;  1B6AF08816F1B8FE CRC64;

  Query Match             28.1%;  Score 1014.5;  DB 10;  Length 708;
  Best Local Similarity   35.6%;  Pred. No. 2e-59;
  Matches  264;  Conservative  122;  Mismatches  269;  Indels   87;  Gaps   17;

Qy         18 NGSHAAP---TKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVELSGYSLT 74
              || | :       ::   | : :     | |      | |: |: :: :      |:| | |||
Db          3 NGHHHSNGLCVETTRDPLNWVAAAESLKG-----SHLDEVKRMVEEFRKPAVKLGGESLT 57

Qy         75 VRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRTED 132
               :   |  :    | |:  :  || |  | |::    :       |||||||| ::  ||:
Db         58 IAQVAAIAARDNAVAVELAETARAGVKASSDWVMESMNKGTDSYGVTTGFGATSHRRTKQ 117

Qy        133 AVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVRLV 192
               :|||  ||     |:           |      |       : ||   | ||:|:|:|  :|:|: :|
Db        118 GGALQKELIRFLNAGI-------FGNGTESNHALPHSATRAAMLVRINTLLQGYSGIRFE 170

Qy        193 VLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFA 252
               :|||| |||| ||| :||||:|:||||| |||||| :|| |: |   |    |  : : |
Db        171 ILEALTKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPNG--EALNA 227

Qy        253 REAISLFGLEAVV-LGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALTVEAMV 311
               || || |:        | ||||| |||||| : :|:::|:|:::|::||: : :  | |
Db        228 GEAFSLAGVSGFFELQPKEGLALVNGTAVGSGLASIALYDANILAVLSEVTSVIFAEVMN 287

Qy        312 GQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDRYPL 371
              |:    |    :  |||:| |   :    |  |: |  |: :::   |       :|||| |
Db        288 GKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHETDPLQKPKQDRYAL 345

Qy        372 RTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTR 431
              |||||:|||  :| :  |    :  | |:  |  ||:||    :   ||||| : : :||: ||
Db        346 RTSPQWLGPQIEVIRTATKMIEREINSVNDTPLIDVSRNKALHGGNFQGTPIGVSMDNTR 405

Qy        432 LALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAAYASELGHLA 490
              ||:| |||| | | :||:|    ||||| |:    :|||:|   || :| :|:| ||| ||
Db        406 LAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGSEIAMASYCSELQFLA 465

Qy        491 NPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFK 550
              ||||  || ||  ||||| |||:|:| ||  :  :| :|::|   ||||||  :|
Db        466 NPVTNHVQSAEQHNQDVNSLGLISSRKTVEALDILKLMSSTYLIALCQAVDLRHLE---- 521

Qy        551 KQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDA--------- 601
                                           | || |  :::  ::| |  :   |:
Db        522 ---------------------ENLRLSVKNTVSQVAKRTLTMGINGELHPSRFCEKDLLR 560

Qy        602 -------FSY----ATGT-----------VVELLSSSPS-ANVTLTAVNAWKVASAEKAI 638
                     |:|    :||            |    |:: |  ||: :    :    |
Db        561 VVDREYVFAYIDDPCSGTYPLMQKLRQVLVDHALNNGESEKNVSTSIFQKIEAFEVELKA 620

Qy        639 SLTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGS 697
              | :|| |      :  | :    |:  || |:||||     |:   |       : |
Db        621 ILPKEVESARIALESGNPAIGNRITECRSYPLYKFIREELGTNYLTGEKVV-----SPGE 675

Qy        698 NVSRIYEAIKDGRINHVLVKML 719
               ::: |: | |  |:| |
Db        676 ECDKVFTALSKGLIVDPLLKCL 697

RESULT 8
Q852S1
ID   Q852S1     PRELIMINARY;      PRT;   715 AA.
AC   Q852S1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   GDCPAL3.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC   Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Kurodagosun;
RA   Ozeki Y., Chikagawa Y., Kimura S., Soh H., Maeda K., Pornsiriwong W.,
RA   Kato M., Akimoto H., Oyanagi M., Fukuda T., Koda T., Itoh Y.,
RA   Yamada A., Ueno H., Takeda J.;
RT   "Putative cis-elements in the promoter region of phenylalanine
RT   ammonia-lyase gene of carrot involved in anthocyanin biosynthetic
RT   pathway.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB089813; BAC56977.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   715 AA;  77802 MW;  8D067795D57565B4 CRC64;

  Query Match             28.1%;  Score 1013;  DB 10;  Length 715;
  Best Local Similarity   36.1%;  Pred. No. 2.6e-59;
  Matches  265;  Conservative  131;  Mismatches  275;  Indels   64;  Gaps   18;

Qy         16 FTNGSHAAPTKSAAGPTSALRRTP----GLDGHAAHQSQLEIVQELLSDPTDDVVELSGY 71
               :||| |     ::  |     :::        |:   |    | || |: ::::      ||:| |
Db          3 YTNGHH----ENGNGVDLCMKKEDPLSWGVAAEALKGSHLEEVKRMVAEYRKPVVKLGGE 58

Qy         72 SLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADT 128
               :||:  |     :||    |:|:    :  ||  |   | ||:    :        ||||||| ::
Db         59 TLTISQVAAISARDDSGVKVELSEAARAGVKASSDWVMESMNKGTDSYGVTTGFGATSHR 118

Qy        129 RTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE---NTLPLEVVRGAMVIRVNSLTRG 185
              ||:    :|||.||      |:                |  |     ||||     | ||::|:|:|: :|
Db        119 RTKQGGALQKELIRFLNAGI---------FGSGAEAGNNTLPHSATRAAMLVRINTLLQG 169

Qy        186 HSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEG 245
               :| :|   :|||:| |||| ||| :||||:|:||||| |||||| :||  |: |
Db        170 YSGIRFEILEAITKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA---VGP 226

Qy        246 TEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALT 303
               |    :     ||   | |:|         | ||||| ||||||| :  ||:: | ::::|::|:: ::
Db        227 TGVTLSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEVMS 286

Qy        304 ALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGI 363
               |:  | | |      :          ||||:|: :   : ||:: |: :::          |
Db        287 AIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEMDPLQK 344

Qy        364 LRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAV 423
               :|||| ||||||:|||  :| :   :      :     | | |:   ||||:||  :  |||||| : :
Db        345 PKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGTPI 404

Qy        424 SISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAAY 482
               :||: ||||:| |||| | | :||:|    | ||||| |:  :|||:|   || :| :|:|
Db        405 GVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASY 464

Qy        483 ASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDL 542
              |||  ||||||  || ||  || |||| |||:|:|:||  ::| |:   :  |      ||:||
```

```
Db        465 CSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAIDL 524

Qy        543 RAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ-------------- 588
              | :| :  |          :  : :: |   |  ||    |    :    | |  :
Db        525 RHLEENLKSTVKNTVSQVAKRVLTMG--VNG-ELHPSRFCEKDLLRVVDREYIFAYIDDP 581

Qy        589 -TTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAIS--LTREVR 645
               : || |   :   :        |:||     ::       |:       |:|: |   :    | :||
Db        582 CSATYPLMQKLRE-------TLVEHALNNGDKERNLSTSIFQKIAAFEDELKALLPKEVE 634

Qy        646 N-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYE 704
              : |        :  |   :     |:  || ||||||       :    |  :   :  |     :::
Db        635 SARAAVESGNPAIPNRIKECRSYPLYKFVREELGT-----EYLTGEKVTSPGEEFDKVFT 689

Qy        705 AIKDGRINHVLVKML 719
              |:  | |    |:: |
Db        690 AMTKGEIIDPLLECL 704
```

```
RESULT 9
Q8H6V6
ID   Q8H6V6      PRELIMINARY;      PRT;   714 AA.
AC   Q8H6V6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
GN   PAL1.
OS   Populus tremuloides (Quaking aspen).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
OX   NCBI_TaxID=3693;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22264006; PubMed=12376645;
RA   Kao Y.Y., Harding S.A., Tsai C.J.;
RT   "Differential expression of two distinct phenylalanine ammonia-lyase
RT   genes in condensed tannin-accumulating and lignifying cells of quaking
RT   aspen.";
RL   Plant Physiol. 130:796-807(2002).
DR   EMBL; AF480619; AAN52279.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   714 AA;  77556 MW;  9BDAADA21CECC35D CRC64;

  Query Match            28.0%;  Score 1010.5;  DB 10;  Length 714;
  Best Local Similarity  35.5%;  Pred. No. 3.8e-59;
  Matches  265;  Conservative  135;  Mismatches  269;  Indels   77;  Gaps   21;

Qy          8 LATTLANGFTNGSHAAPTKSAAGPTSALRRTP---GLDGHAAHQSQLEIVQELLSDPTDD 64
               : |    ||: |||                :   :| |    |:    |    | |: |: ::::
Db          1 METITKNGYQNGS---------SESLCTQRDPLSWGVAAEAMKGSHLDEVKRMVAEYRKP 51

Qy         65 VVELSGYSLTVRDVVG-AARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTG 121
              || |:| :||:   |    |         |:|:   :   | ||  |  | |::     :        |||||
Db         52 VVNLAGQNLTIAQVASIAGHDASNVKVELSESARPRVKASSDWVMDSMDKGTDSYGVTTG 111

Qy        122 FGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE--NTLPLEVVRGAMVIRV 179
              || ::  ||:    :|||  ||          |:              | | |    :|||     | ||::|:
Db        112 FGATSHRRTKQGGALQKELIRFLNAGI---------FGNGTETCHTLPHSATRAAMLVRI 162
```

```
Qy        180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239
              |:| :|:| :|   :|||:|   ||: |||  :||||:|:|||||  ||||||| :|| |:  |
Db        163 NTLLQGYSGIRFEILEAITKLLNNNITPCLPLRGTITASGDLVPLSYIAGLLTGSPNSKA 222

Qy        240 HVLHEGTE-KIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLS 296
                      |    ::: |  ||     |:::      |  ||||  |||||||  : :|:: |  ::::|:
Db        223 ----TGPNGEVLDAVEAFKAAGIDSGFFELQPKEGLALVNGTAVGSGLASMVLFETNVLA 278

Qy        297 LLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVE----HE 352
              :||: ::|:   |  |  |:    |      :     :   |||:|  |    :  :|  ||::        ||
Db        279 VLSELISAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAYMKAAKKLHE 337

Qy        353 EEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQT 412
                :     |         :||||  |||||||:|||  :|  :    :    ::    |  |:    ||||:||        :
Db        338 MDPLQKP-----KQDRYALRTSPQWLGPQIEVIRFSTKSIEREINSVNDNPLIDVSRNKA 392

Qy        413 AHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCL-AAEDPSLNYH 471
              |||||  :  : :||:   |||:|  ||||  |   |   :||:|      |  ||||  |  |:  :|||:|
Db        393 LHGGNFQGTPIGVSMDNVRLAIASIGKLLFAQFSELVNDFYNNGLPSNLTASRNPSLDYG 452

Qy        472 GKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLAS 531
                || :| :|:| |||  :||||||: || ||    || ||||  |||:|:||||  |:| |: :   :
Db        453 FKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSTT 512

Qy        532 HLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTT 591
                |       ||:|||  :| |  :  :          :   : ::   |  || |  :           |      :      | |  :
Db        513 FLVALCQAIDLRHLEENLRSAVKNTVSHVSKRVLTTGANGE---LHPSRFCEKELLKVV- 568

Qy        592 YDLEPRWHDAFSYATG--------------TVVELLSSSPSANVTLTAVNAWKVASAEK 636
               | |      |  | :||                       |         |::   :        |:|     |:|:  |
Db        569 -DRE----DVFAYADDPCSATYPLMQKLRQVLVDHALANGENEKNASTSV-FQKIAAFED 622

Qy        637 AIS--LTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQE 693
                :     |  :||  :   |    :  |   :       |:    ||  ||||||        :      |
Db        623 ELKALLPKEVESARAAYDSGNSAIENKIKECRSYPLYKFVREELGT-----GLLTGENVR 677

Qy        694 TIGSNVSRIYEAIKDGRINHVLVKML 719
                :  |       :::  |: :|:|       ::: |
Db        678 SPGEEFDKVFTAMCEGKIIDPMLECL 703
```

```
RESULT 10
Q7X6M8
ID   Q7X6M8      PRELIMINARY;      PRT;   714 AA.
AC   Q7X6M8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   OSJNBa0073E02.16 protein (OSJNBb0061C13.5 protein).
GN   OSJNBA0073E02.16 OR OSJNBB0061C13.5.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA   Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA   Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA   Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA   Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA   Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA   Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA   Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA   Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL731616; CAE05456.1; -.
DR   EMBL; AL731629; CAE05623.1; -.
SQ   SEQUENCE   714 AA;  76937 MW;  C4C21FF6ECD588FA CRC64;

  Query Match            27.9%;  Score 1007.5;  DB 10;  Length 714;
  Best Local Similarity  36.4%;  Pred. No. 6e-59;
```

```
  Matches  267;  Conservative  130;  Mismatches  263;  Indels   73;  Gaps   21;

Qy         18 NGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVELSGYSLTVRD 77

Db         13 NGSSLCVAKPRADPLNWGKAAEELSG-----SHLDAVKRMVEEYRRPVVTIEGASLTIAQ 67

Qy         78 VVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRTEDAVS 135

Db         68 VAAVASAG-AARVELDESARGRVKASSDWVMNSMMNGTDSYGVTTGFGATSHRRTKEGGA 126

Qy        136 LQKALIEHQLCGVTPTSXSSFSVGRGLE-NTLPLEVVRGAMVIRVNSLTRGHSAVRLVVL 194

Db        127 LQRELIRFLNAG---------AFGNGDDGHVLPAAATRAAMLVRINTLLQGYSGIRFEIL 177

Qy        195 EALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFARE 254

Db        178 ETIATLLNANVTPCLPLRGTITASGDLVPLSYIAGLVTGRPN-SVAVTPDGRK--VDAAE 234

Qy        255 AISLFGLE--AVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALTVEAMVG 312

Db        235 AFKIAGIQHGFFELQPKEGLAMVNGTAVGSGLASMVLFEANVLGVLAEVLSAVFCEVMNG 294

Qy        313 QQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAV------EHEEEVKVKDDEGILRQ 366

Db        295 KP-EYTDHLTHKLKHHPGQIEAAAIMEHILEGSSYMMLAKKLGELDPLMKPK-------Q 346

Qy        367 DRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVSIS 426

Db        347 DRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVS 406

Qy        427 MEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAAYASE 485

Db        407 MDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSE 466

Qy        486 LGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLRAM 545

Db        467 LQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDVLKLMSSTFLIALCQAIDLRHL 526

Qy        546 ELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDAFSYA 605

Db        527 EENVRSAVKGCVTTVARKTLST-----SATGDLHKARFCEKDLLQAIDREA----VFAYA 577

Qy        606 -------------TGTVVE--LLSSSPSANVTLTAVNAWKVASAEK--AISLTREVR-N 646

Db        578 DDPCSANYPLMQKMRAVLIEHALANGEAERNVD-TSVFA-KVATFEEELRVALPREVEAA 635

Qy        647 RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAI 706

Db        636 RAAVENGTAAKANRITECRSYPLYRFVREELGT-----EYLTGEKTRSPGEEVNKVFVAM 690

Qy        707 KDGRINHVLVKML 719

Db        691 NQGKHIDALLECL 703
```

```
RESULT 11
Q84VE0
ID   Q84VE0      PRELIMINARY;      PRT;    671 AA.
AC   Q84VE0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Cooper B., Hutchison D., Park S., Guimil S., Luginbuhl P., Ellero C.,
RA   Goff S., Glazebrook J.;
RT   "Identification of Rice (Oryza sativa) Proteins Linked to the Cyclin-
RT   Mediated Regulation of the Cell Cycle.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY224546; AAO72666.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   671 AA;  72277 MW;  23F93B995CE21E60 CRC64;

  Query Match             27.8%;  Score 1003.5;  DB 10;  Length 671;
  Best Local Similarity   37.3%;  Pred. No. 1e-58;
  Matches  257;  Conservative  138;  Mismatches  239;  Indels   55;  Gaps   18;

Qy         57 LLSDPTDDVVELSGYSLTVRDV--VGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQN- 113

Db          1 MVAQSREAVVKIEGSSLRVGQVAAVSAAKDASGVVVELDEEARPRVKASSEWILNCIAHG 60

Qy        114 -SVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE-NTLPLEVV 171

Db         61 GDIYGVTTGFGGTSHRRTKDGQALQVELLRHLNAGI---------FGNGSDGNSLPSEVS 111

Qy        172 RGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAI 231

Db        112 RAAMLVRINTLLQGYSGIRFEILEAITKLINTGVSPCLPLRGTITASGDLVPLSYIAGLI 171

Qy        232 TGHPDVKVHVLHEGTEKIMFAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSL 289

Db        172 TGRPNAQA-VTVDG--KKVDAAEAFKIAGIQGGFFRLEPKEGLAIVNGTSVGSALAAMVL 228

Qy        290 HDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAV 349

Db        229 YDANVLAVLSEVLSAVFCEVMNGKP-EYTDHLTHKLKHHPGSIEAAAIMEHILAGSAF-M 286

Qy        350 EHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVEN 409

Db        287 PHAQKVNEVDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHR 346

Qy        410 KQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSL 468

Db        347 GKALHGGNFQGTPIGVSMDNTRLAIANIGKLMFAQFSELVNEFYNNGLTSNLAGSRNPSL 406

Qy        469 NYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLL 528

Db        407 DYGFKGTEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLM 466

Qy        529 LASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQ 588

Db        467 SSTYLVALCQAVDLRHLEENLKSAVKNCVTTVAKKVLTTG---PAGGLHSARFSEKAL-- 521

Qy        589 TTTYDLEPRWHDAFSYATG--------------TVVELLSSSPSANVTLTAVNAWKVAS 633

Db        522 LTAIDREA----VYSYADDPCSANYPLMTKIRAVLVEHALANGPAEKDDGSSVFS-KITA 576

Qy        634 AEKAI--SLTREVRNRFWQTPSSQAP-AHAYLSPRTRVLYSFVREELGVQARRGDVFVGV 690

Db        577 FEEELREALPREMEAARVAFETGTAPITNRIKESRSFPLYRFVREELGCV-----YLTGE 631

Qy        691 QQETIGSNVSRIYEAIKDGRINHVLVKML 719

Db        632 KLKSPGEECNKVFLAISERKLIDPMLECL 660
```

RESULT 12

```
Q9XFX5
ID   Q9XFX5      PRELIMINARY;      PRT;   721 AA.
AC   Q9XFX5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine-ammonia lyase (EC 4.3.1.5).
GN   PAL1.
OS   Citrus clementina x Citrus reticulata.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Sapindales; Rutaceae; Citrus.
OX   NCBI_TaxID=93374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Fortune; TISSUE=Flavedo;
RA   Sanchez-Ballesta M.T., Lafuente M.T., Zacarias L., Granell A.;
RT   "Phenylalanine ammonia-lyase gene expression in response to low
RT   temperature in mandarin fruit.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ238753; CAB42793.1; -.
DR   HSSP; P21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   721 AA;  78429 MW;  97DBD6351A5FDAC5 CRC64;

  Query Match             27.8%;  Score 1002.5;  DB 10;  Length 721;
  Best Local Similarity   37.4%;  Pred. No. 1.3e-58;
  Matches  262;  Conservative  124;  Mismatches  253;  Indels   61;  Gaps   18;

Qy         49 SQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKG---RRVRVQNDDEIRARVDKSVD 105
              | || |: ::::   || | | :|||  |   |  |    :|:|:  :  |  | | | |
Db         43 SHLEEVKRMVAEYRKPVVNLGGETLTVAQVAAIATAGDVNAQVKVELSESAREGVKASSD 102

Qy        106 FLKAQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLE 163
              ::  :       ||||||| ::  ||:: :||| ||:      |:         |
Db        103 WVMDSMNKGTDSYGVTTGFGATSHRRTQNGGALQKELIKFLNAGI-------FGNGTKSS 155

Qy        164 NTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSP 223
              :|||    | ||::|||:| :|:| :|  :|:|:| ||| ||| :||||:|:||||| |
Db        156 HTLPHSATRAAMLVRVNTLLQGYSGIRFEILDAITKLLNHSITPCLPLRGTITASGDLVP 215

Qy        224 LSYIAGAITGHPDVKVHVLHEGTE-KIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSA 282
              |||||| :|| |: |       |    :|: |:||   |       | ||||| ||||||| :
Db        216 LSYIAGLLTGRPNSKA----TGPNGEIIDAQEASKQAGFGFFELQPKEGLALVNGTAVGS 271

Qy        283 SMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLL 342
              :|:: | |:: |:||| |:|: | | |: |  |      :  |||:| |   : :|
Db        272 GLASMVLFDANNLALLSEILSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHIL 330

Qy        343 SGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDN 402
              |||: |: :::   |      :|||| ||||||:|||  :| :  |  ::  | |:  ||
Db        331 DGSSY-VKAAKKLHEIDPLQKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDN 389

Qy        403 PLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA 462
              ||:||   :  ||||| : : :|: ||||:|||| |  | | :||:|    ||||  |:
Db        390 PLIDVSRNKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLS 449

Qy        463 -AEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEA 521
               :|||:|  || :|:| :|: |||  |||||| || ||  ||  |||| |||:|:||||
Db        450 GGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEA 509

Qy        522 NDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKA 581
              |:| |: :: |    ||:|||  :| : |     :  : :: |  | |  :    |  :
Db        510 VDILKLMSSTFLVALCQAIDLRHLEENLKHTVKDTVSQVARKVLTVGANGE---LHPSRF 566

Qy        582 LNKRLEQTTTYDLEPRWHDAFSY------ATGTVVELL-----------SSSPSANVTL 623
               | |  :         | |  |:|    || :::  |               :  :|| ::
Db        567 CEKDLLKAA-----DREH-VFAYIDDPCSATYPLMQKLRQVLVEHALNNGENEKNANSSI 620

Qy        624 TAVNAWKVASAEKAIS--LTREVRNRFWQTPSSQAPA--HAYLSPRTRVLYSFVREELGV 679
                    |:|: |: :    |  :|| |   ||   : |   :       |:  ||  ||||||
Db        621 FQ----KIAAFEEELKAVLPKEVENA-RQTVENGNPTIPNRIKECRSYPLYRLVREELGT 675

Qy        680 QARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKML 719
                   :   |  :  :  |   ::: |:  |:|   ::: |
Db        676 -----NFLTGEKVTSPGEKFDKVFTAMCQGKIIDPMLECL 710
```

```
RESULT 13
Q8GZR8
ID   Q8GZR8      PRELIMINARY;      PRT;   713 AA.
AC   Q8GZR8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase2.
OS   Lactuca sativa (Garden lettuce).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;
OC   Lactuca.
OX   NCBI_TaxID=4236;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Campos-Vargas R., Nonogaki H., Suslow T., Saltveit M.E.;
RT   "Characterization of a second phenylalanine ammonia-lyase (PAL) gene
RT   in wounded lettuce leaf tissue.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF411134; AAO13347.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   713 AA;  77657 MW;  87751A2ED7349E61 CRC64;

  Query Match             27.6%;  Score 998;  DB 10;  Length 713;
  Best Local Similarity   36.8%;  Pred. No. 2.6e-58;
  Matches  256;  Conservative  122;  Mismatches  264;  Indels   54;  Gaps   16;

Qy         40 GLDGHAAHQSQLEIVQELLSDPTDDVVELSGYSLTVRDVVG-AARKGRRVRVQNDDEIRA 98
              |:    | |: |: ::::    || | |:|| |   ||     |:|:  :  ||
Db         29 GMAAESLKGSHLDEVKRMVAEFRKPVVRLGGETLTVSQVAAIAASDNAGVKVELSETARA 88

Qy         99 RVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSF 156
               | | |::   :        ||||||| ::  ||::  :|||  ||        |:         |
Db         89 GVKASSDWVMESMNKGTDSYGVTTGFGATSHRRTKEGGALQKELIRFLNAGI-------F 141

Qy        157 SVGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSIS 216
               |      :|||    | ||::|:|:| :|:| :|  :|||:| |||| ||:  :||||:|:
Db        142 GNGTESTHTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNHNVTPFLPLRGTIT 201

Qy        217 ASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEA--VVLGPKEGLGL 274
              ||||| |||||||| :|| :      | ::: |:| : |:|      |||||: |
Db        202 ASGDLVPLSYIAGLLTGRANSKA---VGPTGEVLNAEKAFAEAGVEGGFFELQPKEGLAL 258
```

```
Qy        275 VNGTAVSASMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEV 334
Db        259 VNGTAVGSGMASMVLFDANVLALLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEA 317

Qy        335 ARNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSL 394
Db        318 AAIMEYILDGSDY-VKAAQKVHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRSSTKMIER 376

Qy        395 ENNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLNAAMN 454
Db        377 EINSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYN 436

Qy        455 RGLPSCLA-AEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVNSLALI 513
Db        437 NGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLI 496

Qy        514 SARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA 573
Db        497 SARKTAESVEILKLMSTTYLVALCQSIDLRHLEENLKSTVKNTVSLVAKKILTTG--VNG 554

Qy        574 LALEVKKALNKRLEQ---------------TTTYDLEPRWHDAFSYATGTVVELLSSSPS 618
Db        555 -ELHPSRFCEKDLLRVVDREYVFAYIDDACSATYPLMQKLRQVI------VDHALNNEND 607

Qy        619 ANVTLTAVNAWKVASAEKAIS--LTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVRE 675
Db        608 AGTSIFQ----KISEFEEELKAVLPKEVEGVRSAYESSTLTIPNRIKECRSYPLYRFVRE 663

Qy        676 ELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRI 711
Db        664 ELGT-----GFLTGEEVTSPGEEFDKVFTALCKGHI 694
```

```
RESULT 14
Q8VXG7
ID   Q8VXG7      PRELIMINARY;      PRT;   703 AA.
AC   Q8VXG7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Corso;
RX   MEDLINE=97161122;
RA   Rosler J., Krekel F., Amrhein N., Schmid J.;
RT   "Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase
RT   activity.";
RL   Plant Physiol. 113:175-179(1997).
DR   EMBL; L77912; AAL40137.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   703 AA;  74926 MW;  11374FD68516971E CRC64;

  Query Match             27.6%;  Score 997.5;  DB 10;  Length 703;
  Best Local Similarity   35.6%;  Pred. No. 2.8e-58;
  Matches  262;  Conservative  133;  Mismatches  247;  Indels   93;  Gaps   21;
```

```
Qy         29 AGPTSALRRTPGLDGHAAHQ---SQLEIVQELLSDPTDDVVELSGYSLTVRDV--VGAAR 83
Db          2 AGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVASAK 61

Qy         84 KGRRVRVQNDDEIRARVDKSVDFLKAQLQN--SVYGVTTGFGGSADTRTEDAVSLQKALI 141
Db         62 DASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSHRRTKDGPALQVELL 121

Qy        142 EHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFL 201
Db        122 RHLNAGIFGTGSDG-------HTLPSEVTRAAMLVRINTLLQGYSGIRFEILEAITKLL 173

Qy        202 NHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGL 261
Db        174 NTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQA-VTVDGRK--VDAAEAFKIAGI 230

Qy        262 EA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAP 319
Db        231 EGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKP-EYTD 289

Qy        320 FIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGIL--RQDRYPLRTSPQF 377
Db        290 HLTHKLKHHPGSIEAAAIMEHILDGSSFM---KQAKKVNELDPLLKPKQDRYALRTSPQW 346

Qy        378 LGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALI 437
Db        347 LGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANI 406

Qy        438 GKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAAYASELGHLANPVTTF 496
Db        407 GKLMFAQFSELVNEFYNNGLTSNLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNH 466

Qy        497 VQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPL 556
Db        467 VQSADEHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAVDLRHLEENIKAS---- 522

Qy        557 LPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDAFSYATGTVVELLSSS 616
Db        523 VKNTVTQVAKKVLTMNPSG-ELSSARFSEKELISAIDREA----VFTYA----------E 567

Qy        617 PSANVTLTAVNAWKVASAEKAIS-----------LTREVRNRFWQTPSSQAP------- 657
Db        568 DAASASLPLMQKLRAVLVDHALSSGERGAGALRVLQDHQVRG---GAPRGAAPGGGGRPR 624

Qy        658 ---------AHAYLSPRTRVLYSFVREELGVQARRGD-----------VFVGVQQETIGS 697
Db        625 GVAEGTAPVANRIADSRSFPLYRFVREELGCVFLTGERLKSPGEECNKVFVGISQ---GK 681

Qy        698 NVSRIYEAIK--DGR 710
Db        682 LVDPMLECLKEWDGK 696
```

```
RESULT 15
Q9ATN7
ID   Q9ATN7      PRELIMINARY;      PRT;   716 AA.
AC   Q9ATN7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
GN   PAL.
OS   Agastache rugosa.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Lamiales; Lamiaceae; Nepetoideae; Mentheae; Agastache.
OX   NCBI_TaxID=39271;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim J.-B., Kim J.-B., Cho K.-J., Woon U., Ha S.-H.;
RT   "Molecular Cloning and Characterization of Phenylalanine Ammonia-
RT   lyase(PAL) Related to the Biosynthesis of Rosmarinic acid in Agastache
```

```
RT   rugosa.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF326116; AAK15640.1; -.
DR   HSSP; P21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   716 AA;  77487 MW;  9903F33B8CFD2958 CRC64;

  Query Match             27.6%;  Score 997;  DB 10;  Length 716;
  Best Local Similarity   35.2%;  Pred. No. 3.1e-58;
  Matches  262;  Conservative  126;  Mismatches  271;  Indels   86;  Gaps   16;

Qy         13 ANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSDPTDDVVELSGYS 72

Db          9 SNGHNNGANGFCVKQ-NDPLNWAAAAESLKG-----SHLEEVKRMVEEFRKPVVKLGGET 62

Qy         73 LTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGFGGSADTRT 130

Db         63 LTISQVAAIAAKDNAVAVELAESARAGVKASSDWVMDSMSKGTDSYGVTTGFGATSHRRT 122

Qy        131 EDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVNSLTRGHSAVR 190

Db        123 KQGGALQKELIRFLNAGIFGNATES-------NHTLPHTATRAAMLVRINTLLQGYSGIR 175

Qy        191 LVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKVHVLHEGTEKIM 250

Db        176 FEILEAITKFLNQNVTPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPAG--EPL 232

Qy        251 FAREAISLFGLEA--VVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLSQALTALTVE 308

Db        233 TAEQAFKLAGVTGGFFDLQPKEGLALVNGTAVGSGLASIALFDANVLAVLSVVMSPVFAE 292

Qy        309 AMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDR 368

Db        293 VMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSGY-VKAAQKLHEIDPLQKPKQDR 350

Qy        369 YPLRTSPQFLGPLVEDMMHAYSTLSLENNTTTDNPLLDVENKQTAHGGNFQASAVSISME 428

Db        351 YALRTSPQWLGPQIEVIRTATKMIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMD 410

Qy        429 KTRLALALIGKLNFTQCTELLNAAMNRGLPSCLA-AEDPSLNYHGKGLDIHIAAYASELG 487

Db        411 NTRLAIASIGKLLFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGSEIAMASYCSELQ 470

Qy        488 HLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQAVDLRAMEL 547

Db        471 FLVNPVTNHVQSAEQHNQDVNSLGLISSRKTVEALDILKLMSSTYLVALCQAVDLRHVEE 530

Qy        548 DFKKQFDPLLPTLLQQHLGTGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDA------ 601

Db        531 NMK-------------------------LAVKNTVSQVAKRTLTMGVNGELHPSRFCEKE 565

Qy        602 ----------FSY------ATGTVVE----------LLSSSPSANVTLTAVNAWKVASAE 635

Db        566 LIRVVDREYVFTYIDDPCLATYPLMQKLRQVLVDHALKNGESEKNVSTSIFHKIEAFEEE 625

Qy        636 KAISLTREVRN-RFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQET 694

Db        626 LKALLPKEVESARIALESGSPAVANRIEECRSFPLYKFIREELGTGFLTGEKAV-----S 680

Qy        695 IGSNVSRIYEAIKDGRINHVLVKML 719

Db        681 PGEECEKVFAALSNGLIIDPLLECL 705

Search completed: September  9, 2004, 10:07:42
Job time : 91.6349 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:         September 11, 2004, 10:06:53 ; Search time 15236.4 Seconds
                                       (without alignments)
                                      7040.653 Million cell updates/sec

Title:          US-09-939-408A-20
Perfect score:  2475
Sequence:       1 atggcccctccntcgactc..........nnnnnnnnnnnnnnnnnnn 2475

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:      6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2128	86.0	2475	6	AR199606	AR199606 Sequence
	2	2128	86.0	2475	6	AX366868	AX366868 Sequence
	3	1699.4	68.7	2439	6	AR199605	AR199605 Sequence
	4	1699.4	68.7	2439	6	AX366866	AX366866 Sequence
	5	1677.4	67.8	2331	6	E01783	E01783 DNA sequenc
	6	1677.4	67.8	2331	6	E01784	E01784 DNA sequenc
	7	1677.4	67.8	2331	6	E01785	E01785 DNA sequenc
	8	1676.4	67.7	2315	8	RTPALGNA	X12702 R.toruloide
	9	1629.4	65.8	2419	6	AR199602	AR199602 Sequence
	10	1629	65.8	2419	6	AX366860	AX366860 Sequence
	11	1626.4	65.7	2151	6	E01759	E01759 DNA encodin
	12	1626.4	65.7	2151	6	AR205032	AR205032 Sequence
	13	1626.4	65.7	2151	6	AR282639	AR282639 Sequence
	14	1626.4	65.7	2151	6	AX082544	AX082544 Sequence
	15	1626.4	65.7	2151	6	AX370659	AX370659 Sequence
	16	1623.2	65.6	2151	6	E01543	E01543 L-phenylala
	17	1623.2	65.6	2151	6	AR205031	AR205031 Sequence
	18	1623.2	65.6	2151	6	AR282638	AR282638 Sequence
	19	1623.2	65.6	2151	6	AX082542	AX082542 Sequence
	20	1623.2	65.6	2151	6	AX370657	AX370657 Sequence
	21	1610.4	65.1	2061	6	AR282658	AR282658 Sequence
	22	1610.4	65.1	2061	6	AX370681	AX370681 Sequence
	23	1598.6	64.6	2787	6	AR199604	AR199604 Sequence
	24	1598.6	64.6	2787	6	AX366864	AX366864 Sequence
	25	1258	50.8	3340	8	RTPAL2	X51513 Rhodosporid
	26	1213.6	49.0	2986	8	RHDPAL	M18261 R. toruloid
	27	1156.6	46.7	3586	8	RRPAL	X13094 Rhodotorula
	28	1124.4	45.4	2741	6	AX366876	AX366876 Sequence
	29	461.8	18.7	2488	8	MZEPAL	L77912 Zea mays ph
	30	450.2	18.2	2504	8	AK102817	AK102817 Oryza sat
	31	447	18.1	2280	6	AX654420	AX654420 Sequence
	32	442.2	17.9	2124	6	AX653389	AX653389 Sequence
	33	442.2	17.9	2124	6	AX660330	AX660330 Sequence
	34	442.2	17.9	2510	8	AK100346	AK100346 Oryza sat
	35	440.4	17.8	2016	8	AY224546	AY224546 Oryza sat
	36	440	17.8	2282	6	AX660254	AX660254 Sequence
	37	436.8	17.6	2126	6	AX660228	AX660228 Sequence
c	38	429.6	17.4	139226	8	OSJN00274	AL731629 Oryza sat
c	39	429.6	17.4	139488	8	OSJN00262	AL731616 Oryza sat
	40	426.4	17.2	149953	2	AP005000	AP005000 Oryza sat
c	41	426.4	17.2	154494	2	AP005070	AP005070 Oryza sat
c	42	426.4	17.2	165957	8	AP006523	AP006523 Oryza sat
c	43	422.6	17.1	170186	2	AC135419	AC135419 Oryza sat
c	44	422.6	17.1	177374	8	AC135429	AC135429 Oryza sat
	45	421.8	17.0	4412	8	OSPAL	X16099 Rice gene f

ALIGNMENTS

```
RESULT 1
AR199606
LOCUS       AR199606                2475 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 20 from patent US 6355468.
ACCESSION   AR199606
VERSION     AR199606.1  GI:20249680
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2475)
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: US 6355468-A 20 12-MAR-2002;
```

```
FEATURES             Location/Qualifiers
     source          1..2475
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             86.0%;  Score 2128;  DB 6;  Length 2475;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-224;
  Matches 2408;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
```

```
Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy       2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNN 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNN 2400

Qy       2401 NNNANAAA 2408
              ||||||||
Db       2401 NNNANAAA 2408


RESULT 2
AX366868
LOCUS       AX366868                2475 bp    DNA     linear   PAT 15-FEB-2002
DEFINITION  Sequence 20 from Patent WO0208402.
ACCESSION   AX366868
VERSION     AX366868.1  GI:18698157
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: WO 0208402-A 20 31-JAN-2002;
            PCBU Services, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2475
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Consensus Sequence of SEQ ID NOs: 12, 16, and 18"
ORIGIN

  Query Match             86.0%;  Score 2128;  DB 6;  Length 2475;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-224;
  Matches 2408;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy        1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy        1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy        1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy        1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy        1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy        1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy        1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy        1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy        1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy        1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy        1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy        2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy        2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy        2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy        2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy        2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy        2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy        2401 NNNANAAA 2408
                ||||||||
Db        2401 NNNANAAA 2408
```

```
RESULT 3
AR199605
LOCUS       AR199605                2439 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 18 from patent US 6355468.
ACCESSION   AR199605
VERSION     AR199605.1  GI:20249679
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2439)
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: US 6355468-A 18 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..2439
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             68.7%;  Score 1699.4;  DB 6;  Length 2439;
  Best Local Similarity   79.4%;  Pred. No. 5.1e-177;
  Matches 1913;  Conservative    0;  Mismatches  459;  Indels   36;  Gaps    6;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
              ||||| ||||  |||||||||| || |||| |||| ||||| ||  |  ||   |
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              || || |               | ||||| || |||  |||      || | ||| |
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              ||||| ||| || || |||||    |||||| ||||||||| ||| |||| ||||||
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
               || ||||    ||       ||||||||||| ||| | |||| ||||| |||||
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
               ||  |||||  | ||||||  |  |||||||||    ||||| |||||||||||| ||
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
              ||||| |||||     |||||||| | ||   | || |||  ||  || ||||||| |||
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
              ||||||| || |||| || || ||||||| || ||||| || ||||||||||||||
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
              || ||||||||||||||||||||||||| ||||| ||||| |||| ||| ||| ||||
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
              || ||||| |||||||||||||||| |||||||| || |||||||||||||| ||||||
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
              |||||      | || ||||||||||||||| |||||||||||||||||||||||||||||
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
```

```
Db      637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy      721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db      697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db      757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db      817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy      901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db      877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db      937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db      997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy     1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db     1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy     1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db     1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy     1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db     1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy     1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db     1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy     1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db     1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy     1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db     1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db     1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db     1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db     1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db     1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740
Db     1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db     1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db     1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db     1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db     1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db     1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db     2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db     2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db     2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy     2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db     2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy     2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db     2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCCGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy     2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db     2305 AGTACCCATCCACGCATCCGATAAATCAGGGAGAGAATCTACGCTTGCGGGAGCTTCTTG 2364

Qy     2401 NNNANAAA 2408
Db     2365 CGCATAAA 2372
```

```
RESULT 4
AX366866
LOCUS       AX366866               2439 bp    DNA     linear   PAT 15-FEB-2002
DEFINITION  Sequence 18 from Patent WO0208402.
ACCESSION   AX366866
VERSION     AX366866.1  GI:18698155
KEYWORDS    .
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: WO 0208402-A 18 31-JAN-2002;
            PCBU Services, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2439
                     /organism="Rhodosporidium toruloides"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5286"
     CDS             1..2151
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD23831.1"
```

```
                     /db_xref="GI:18698156"
                     /db_xref="REMTREMBL:CAD23831"
                     /translation="MAPSLDSISHSFANGVASAKQAVNGASTNLAVAGSHLPTTQVTQ
                     VDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLR
                     SQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSL
                     PLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL
                     SYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSA
                     SMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRK
                     LLEGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQ
                     STTDNPLIDVENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNR
                     GLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALI
                     SARRTTESNDVLSLLATHLYCVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTG
                     SNLRDELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNA
                     WKVAAAESAISLTRQVRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDV
                     FLGKQEVTIGSNVSKIYEAIKSGRINNVLLKMLA"
ORIGIN

  Query Match             68.7%;  Score 1699.4;  DB 6;  Length 2439;
  Best Local Similarity   79.4%;  Pred. No. 5.1e-177;
  Matches 1913;  Conservative    0;  Mismatches  459;  Indels   36;  Gaps    6;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
```

```
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Db       2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Db       2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCCGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Db       2305 AGTACCCATCCACGCATCCGATAAATCAGGGAGAGAATCTACGCTTGCGGGAGCTTCTTG 2364

Qy       2401 NNNANAAA 2408

Db       2365 CGCATAAA 2372
```

```
RESULT 5
E01783
LOCUS       E01783                   2331 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for L-phenylalanine ammonialyase.
ACCESSION   E01783
VERSION     E01783.1  GI:2170036
KEYWORDS    JP 1988317086-A/1.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2331)
  AUTHORS   Fukuhara,N., Yoshino,S., Watanabe,M., Nakajima,Y. and Makiguchi,N.
  TITLE     RECOMBINANT PLASMID FOR MANIFESTING L-PHENYLALANINE-AMMONIA-LYASE
            AND TRANSFORMANT CONTAINING SAID PLASMID
  JOURNAL   Patent: JP 1988317086-A 1 26-DEC-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium toruloides
            PN   JP 1988317086-A/1
            PD   26-DEC-1988
            PF   18-JUN-1987 JP 1987152357
            PI   FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
            NAKAJIMA YOSHIYUKI,
            PI   MAKIGUCHI NOBUYOSHI
            PC   C12N15/00,C12N1/20//C12N9/88,(C12N1/20,C12R1:01),(C12N9/88, PC
              C12R1:01);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=IFO 559 (ATCC 10788);
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1..2151
            FT                   /product='L-phenylalanine ammonialyase' FT
            3'UTR           2152..2331.
FEATURES             Location/Qualifiers
     source          1..2331
                     /organism="Rhodosporidium toruloides"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5286"
ORIGIN

  Query Match             67.8%;  Score 1677.4;  DB 6;  Length 2331;
  Best Local Similarity   80.4%;  Pred. No. 1.3e-174;
  Matches 1891;  Conservative    0;  Mismatches  425;  Indels   36;  Gaps    6;
```

```
Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGAATCCTTCNGCCTC 480

Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876
```

```
Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db    997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy   2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db   2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy   2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db   2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy   2341 NGTNNNCANNNA 2352
Db   2305 AGTACCCATCCA 2316
```

```
RESULT 6
E01784
LOCUS       E01784                 2331 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for L-phenylalanine ammonialyase.
ACCESSION   E01784
VERSION     E01784.1  GI:2170037
KEYWORDS    JP 1988317087-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2331)
  AUTHORS   Fukuhara,N., Yoshino,S., Watanabe,M., Nakajima,Y. and Makiguchi,N.
  TITLE     CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF
            ADVENTITIOUS GENE PRODUCT USING SAID CONTROL
  JOURNAL   Patent: JP 1988317087-A 1 26-DEC-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium
            PN   JP 1988317087-A/1
            PD   26-DEC-1988
            PF   18-JUN-1987 JP 1987152358
            PI   FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
            NAKAJIMA YOSHIYUKI,
            PI   MAKIGUCHI NOBUYOSHI
            PC   C12N15/00,C12N9/88,(C12N9/88,C12R1:01);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=IFO 559 (ATCC 10788);
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1. .2151
            FT                   /product='L-phenylalanine ammonialyase' FT
            3'UTR         2152. .2331.
FEATURES             Location/Qualifiers
     source          1. .2331
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

  Query Match             67.8%;  Score 1677.4;  DB 6;  Length 2331;
  Best Local Similarity   80.4%;  Pred. No. 1.3e-174;
  Matches 1891;  Conservative    0;  Mismatches  425;  Indels   36;  Gaps    6;
```

```
Qy      1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db      1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy     61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db     61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy    121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db    101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy    181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db    161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy    241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db    218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy    301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db    277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy    361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db    337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy    421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db    397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy    481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db    457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy    541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db    517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy    601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db    577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy    661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db    637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy    721 GTNCACGTNNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db    697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy    781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db    757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy    841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db    817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db    997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
```

```
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Db       2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Db       2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy       2341 NGTNNNCANNNA 2352

Db       2305 AGTACCCATCCA 2316
```

```
RESULT 7
E01785
LOCUS       E01785                  2331 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA sequence coding for L-plenylalanine ammonialyase.
ACCESSION   E01785
VERSION     E01785.1  GI:2170038
KEYWORDS    JP 1988317088-A/1.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2331)
  AUTHORS   Fukuhara,N., Yoshino,S., Watanabe,M., Yamamoto,K., Suzuki,M. and
            Nakajima,Y.
  TITLE     CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF
            ADVENTITIOUS GENE PRODUCT USING SAID CONTROL
  JOURNAL   Patent: JP 1988317088-A 1 26-DEC-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium toruloides
            PN   JP 1988317088-A/1
            PD   26-DEC-1988
            PF   18-JUN-1987 JP 1987152359
            PI   FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
            YAMAMOTO KAORU,
            PI   SUZUKI MAKI, NAKAJIMA YOSHIYUKI
            PC   C12N15/00,C12N9/88,(C12N9/88,C12R1:01);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=IFO 559 (ATCC 10788);
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             1..2151
            FT                   /product='L-phenylalanine ammonialyase' FT
            3'UTR           2152..2331.
FEATURES             Location/Qualifiers
     source          1..2331
                     /organism="Rhodosporidium toruloides"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5286"
ORIGIN

  Query Match            67.8%;  Score 1677.4;  DB 6;  Length 2331;
  Best Local Similarity  80.4%;  Pred. No. 1.3e-174;
  Matches 1891;  Conservative    0;  Mismatches  425;  Indels   36;  Gaps    6;
```

```
Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236
```

```
Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db       2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db       2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy       2341 NGTNNNCANNNA 2352
Db       2305 AGTACCCATCCA 2316
```

```
RESULT 8
RTPALGNA
LOCUS       RTPALGNA                2315 bp    mRNA    linear   PLN 23-MAR-1993
DEFINITION  R.toruloides mRNA for L-phenylalanine ammonia-lyase (PAL).
ACCESSION   X12702
VERSION     X12702.1  GI:288350
KEYWORDS    L-phenylalanine ammonia-lyase; PAL gene.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2315)
  AUTHORS   Unknown.
  JOURNAL   Unpublished
COMMENT     (Patented) EP_0_260_919_A1.
FEATURES             Location/Qualifiers
     source          1..2315
                     /organism="Rhodosporidium toruloides"
                     /mol_type="mRNA"
                     /strain="IFO 559"
                     /db_xref="taxon:5286"
                     /clone="pSW13"
     gene            1..2315
                     /gene="PAL"
     CDS             1..2151
                     /gene="PAL"
                     /codon_start=1
                     /product="L-phenylalanine ammonia-lyase"
                     /protein_id="CAA31209.1"
                     /db_xref="GI:288351"
                     /db_xref="GOA:P11544"
                     /db_xref="SWISS-PROT:P11544"
                     /translation="MAPSLDSISHSFANGVASAKQAVNGASTNLAVAGSHLPTTQVTQ
                     VDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLR
                     SQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSL
                     PLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL
                     SYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSA
                     SMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRK
                     LLEGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQ
                     STTDNPLIDVENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNR
                     GLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALI
                     SARRTTESNDVLSLLLATHLYCVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTG
                     SNLRDELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNA
                     WKVAAAESAISLTRQVRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDV
                     FLGKQEVTIGSNVSKIYEAIKSGRINNVLLKMLA"
     polyA_site      2315
                     /gene="PAL"
ORIGIN

  Query Match             67.7%;  Score 1676.4;  DB 8;  Length 2315;
  Best Local Similarity   80.5%;  Pred. No. 1.7e-174;
  Matches 1890;  Conservative    0;  Mismatches  422;  Indels   36;  Gaps    6;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217
```

```
Qy    241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db    218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy    301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db    277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy    361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db    337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy    421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db    397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy    481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db    457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy    541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db    517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy    601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db    577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy    661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db    637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy    721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db    697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy    781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db    757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy    841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db    817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db    997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy   2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db   2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy   2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db   2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy   2341 NGTNNNCA 2348
Db   2305 AGTACCCA 2312
```

```
RESULT 9
AR199602
LOCUS       AR199602                2419 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 12 from patent US 6355468.
ACCESSION   AR199602
VERSION     AR199602.1  GI:20249676
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2419)
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: US 6355468-A 12 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1..2419
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             65.8%;  Score 1629.4;  DB 6;  Length 2419;
  Best Local Similarity   78.0%;  Pred. No. 2.3e-169;
  Matches 1879;  Conservative    0;  Mismatches  504;  Indels   25;  Gaps    8;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGTCTCGTCCTTCAGCGTC 510

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db       1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740
Db       1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC-----------AACGCA 1755

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
```

```
Db       1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Db       1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db       1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db       1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db       2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db       2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Db       2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Db       2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Db       2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352

Qy       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Db       2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411

Qy       2401 NNNANAAA 2408

Db       2412 AAAAAAAA 2419
```

```
RESULT 10
AX366860
LOCUS       AX366860                2419 bp    DNA     linear   PAT 15-FEB-2002
DEFINITION  Sequence 12 from Patent WO0208402.
ACCESSION   AX366860
VERSION     AX366860.1  GI:18698149
KEYWORDS    .
SOURCE      Rhodotorula graminis
  ORGANISM  Rhodotorula graminis
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
            Rhodotorula.
REFERENCE   1
  AUTHORS   Yoshida,R.K. and Kootstra,A.B.
  TITLE     Phenylalanine ammonia lyase polypeptide and polynucleotide
            sequences and methods of obtaining and using same
  JOURNAL   Patent: WO 0208402-A 12 31-JAN-2002;
            PCBU Services, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2419
                     /organism="Rhodotorula graminis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:29898"
     CDS             37..2199
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD23828.1"
                     /db_xref="GI:18698150"
                     /db_xref="REMTREMBL:CAD23828"
                     /translation="MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGH
                     AAHQSQLEIVQELLSDPTDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDK
                     SVDFLKAQLQNSVYGVTTGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGR
                     GLENTLPLEVVRGAMVIRVNSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISAS
                     GDLSPLSYIAGAITGHPDVKVHVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVN
                     GTAVSASMATLSLHDSHMLSLLSQALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEV
                     ARNIRTLLSGSSFAVEHEEEVKVKDDEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTL
                     SLENNTTTDNPLLDVENKQTAHGGNFQASAVSISMEKTRLALALIGKLNFTQCTELLN
                     AAMNRGLPSCLAAEDPSLNYHGKGLDIHIAAYASELGHLANPVTTFVQPAEMGNQAVN
                     SLALISARRTAEANDVLSLLLASHLYCTLQAVDLRAMELDFKKQFDPLLPTLLQQHLG
                     TGLDVNALALEVKKALNKRLEQTTTYDLEPRWHDAFSYATGTVVELLSSSPSANVTLT
                     AVNAWKVASAEKAISLTREVRNRFWQTPSSQAPAHAYLSPRTRVLYSFVREELGVQAR
                     RGDVFVGVQQETIGSNVSRIYEAIKDGRINHVLVKMLA"
     unsure          493..495
                     /note="Other information: Xaa = Val or Ala"
ORIGIN

  Query Match             65.8%;  Score 1629;  DB 6;  Length 2419;
  Best Local Similarity   78.0%;  Pred. No. 2.5e-169;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;
```

```
Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Db         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Db         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Db        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Db        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Db        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Db        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Db        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Db        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Db        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Db        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Db        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810
```

```
Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db      811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db      871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930

Qy      901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db      931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990

Qy      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db      991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050

Qy     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db     1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110

Qy     1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db     1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170

Qy     1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db     1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227

Qy     1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db     1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287

Qy     1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db     1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347

Qy     1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db     1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407

Qy     1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db     1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467

Qy     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db     1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527

Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db     1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db     1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db     1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db     1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC------------AACGCA 1755

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db     1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db     1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db     1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db     1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993

Qy     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db     1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db     2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db     2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db     2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232

Qy     2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db     2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292

Qy     2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db     2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352

Qy     2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db     2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411

Qy     2401 NNNANAAA 2408
Db     2412 AAAAAAAA 2419
```

```
RESULT 11
E01759
LOCUS       E01759                  2151 bp    RNA     linear   PAT 29-SEP-1997
DEFINITION  DNA encoding L-phenylanine anmonia-lyase.
ACCESSION   E01759
VERSION     E01759.1  GI:2170012
KEYWORDS    JP 1988291583-A/1.
SOURCE      Rhodosporidium toruloides
  ORGANISM  Rhodosporidium toruloides
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
REFERENCE   1  (bases 1 to 2151)
  AUTHORS   Fukuhara,N., Yoshino,S., Yamamoto,K., Nakajima,Y., Iwata,T.,
            Watanabe,M., Suzuki,M. and Makiguchi,N.
  TITLE     AMINO ACID SEQUENCE OF L-PHENYLALANINE AMMONIA LYASE, ITS
            STRUCTURAL GENE, NOVEL BASE SEQUENCE CONTAINING SAME, TRANSFORMANT
            PRODUCED THEREWITH AND PRODUCTION OF L-PHENYLALANINE USING SAME
  JOURNAL   Patent: JP 1988291583-A 1 29-NOV-1988;
            MITSUI TOATSU CHEM INC
COMMENT     OS   Rhodosporidium toruloides
            PN   JP 1988291583-A/1
            PD   29-NOV-1988
            PF   22-MAY-1987 JP 1987123950
            PI   FUKUHARA NOBUHIRO, YOSHINO SADAO, YAMAMOTO KAORU, PI
            NAKAJIMA YOSHIYUKI,
            PI   IWATA TOMOYUKI, WATANABE MIDORI, SUZUKI MAKI, PI   MAKIGUCHI
            NOBUYOSHI
            PC   C12N9/88,C12N1/16,C12N1/18,C12N1/20,C12N15/00,C12P13/22, PC
            (C12P13/22,
            PC   C12R1:19),(C12P13/22,C12R1:865);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
```

```
          CC    anti-sense: No;
          CC    *source: strain=IFO 559;
          CC    *source: clone=pSW101;
          FH    Key             Location/Qualifiers
          FH
          FT    CDS             1. .2151
          FT                    /product='L-phenylananine anmonia-lyase' FT
          mat_peptide     1. .2148
          FT                    /product='L-phenylanine anmonia-lyase'.
FEATURES             Location/Qualifiers
     source          1. .2151
                     /organism="Rhodosporidium toruloides"
                     /mol_type="genomic RNA"
                     /db_xref="taxon:5286"
ORIGIN

  Query Match            65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity  83.7%;  Pred. No. 5e-169;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
```

```
Db      1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy      1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db      1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy      1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db      1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy      2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db      2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy      2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db      2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy      2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186

Db      2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151


RESULT 12
AR205032
LOCUS       AR205032                2151 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 9 from patent US 6368837.
ACCESSION   AR205032
VERSION     AR205032.1  GI:21502510
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2151)
  AUTHORS   Gatenby,A.A., Sariaslani,S., Tang,X.-S., Qi,W.Wei. and Vannelli,T.
  TITLE     Bioproduction of para-hydroxycinnamic acid
  JOURNAL   Patent: US 6368837-A 9 09-APR-2002;
FEATURES             Location/Qualifiers
     source          1..2151
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 5e-169;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy         1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Db         1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy        61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Db        61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy       121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Db       101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy       181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Db       161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy       241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Db       218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy       301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Db       277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy       361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Db       337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy       421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Db       397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy       481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Db       457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy       541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Db       517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy       601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Db       577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy       661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db       637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy       721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db       697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy       781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db       757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy       841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db       817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy       901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Db       877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy       961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Db       937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy      1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Db       997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy      1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Db      1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy      1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Db      1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy      1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Db      1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy      1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Db      1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy      1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Db      1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy      1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Db      1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy      1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Db      1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476
```

```
Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

```
RESULT 13
AR282639
LOCUS       AR282639                 2151 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 9 from patent US 6521748.
ACCESSION   AR282639
VERSION     AR282639.1  GI:29719238
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2151)
  AUTHORS   Tang,X.-S.
  TITLE     Polynucleotide encoding a mutant Rhodotorula glutinis tyrosine
            ammonia lyase polypeptide
  JOURNAL   Patent: US 6521748-A 9 18-FEB-2003;
FEATURES             Location/Qualifiers
     source          1..2151
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match            65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity  83.7%;  Pred. No. 5e-169;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;
```

```
Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056
```

```
Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

```
RESULT 14
AX082544
LOCUS       AX082544                2151 bp    DNA     linear   PAT 28-FEB-2001
DEFINITION  Sequence 9 from Patent WO0111071.
ACCESSION   AX082544
VERSION     AX082544.1  GI:13184657
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Tang,X.S., Vannelli,T.M., Qi,W.W., Sariaslani,S. and Gatenby,A.A.
  TITLE     Bioproduction of para-hydroxycinnamic acid
  JOURNAL   Patent: WO 0111071-A 9 15-FEB-2001;
            E.I. DU PONT DE NEMOURS & COMPANY INCORPORATED (US)
FEATURES             Location/Qualifiers
     source          1..2151
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="mutant from Rhodotorula glutinis"
ORIGIN

  Query Match             65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 5e-169;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
```

```
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186

Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

```
RESULT 15
AX370659
LOCUS       AX370659                2151 bp    DNA     linear   PAT 01-MAR-2002
DEFINITION  Sequence 9 from Patent WO0210407.
ACCESSION   AX370659
VERSION     AX370659.1  GI:19168824
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Gatenby,A.A., Sariaslani,F.S., Tang,X.S., Qi,W.W. and Vannelli,T.
  TITLE     Bioproduction of para-hydroxycinnamic acid
  JOURNAL   Patent: WO 0210407-A 9 07-FEB-2002;
            E.I. DUPONT DE NEMOURS AND COMPANY (US)
FEATURES             Location/Qualifiers
     source          1..2151
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     CDS             1..2151
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD26694.1"
                     /db_xref="GI:19168825"
                     /db_xref="REMTREMBL:CAD26694"
                     /translation="MAPSLDSISHSFANGVASAKQAVNGASTNLAVAGSHLPTTQVTQ
                     VDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLR
                     SQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSL
                     PLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL
                     SYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSA
                     SMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRK
                     LLEGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQ
                     STTDNPLIDVENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNR
                     GLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALI
                     SARRTTESNDVLSLLLATHLYCVLQAIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTG
                     SNLRDELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNA
```

```
                    WKVAAAESAISLTRQVRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDV
                    FLGKQEVTIGSNVSKIYEAIKSGRINNVLLKMLA"
     misc_feature   1. .2151
                    /note="mutant from Rhodotorula glutinis"
ORIGIN

  Query Match            65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity  83.7%;  Pred. No. 5e-169;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005
```

```
Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
              ||||||||||||| |||||||||||||||||||| |||||||||||||||||||| ||||
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
              ||||||||||| | |||||||||   ||||||||||||||||||   ||| | |||||| 
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
              |||||||| ||||||||||||| |||
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

Search completed: September 11, 2004, 23:12:51
Job time : 15249.4 secs

```
OM nucleic - nucleic search, using sw model

Run on:          September 11, 2004, 10:03:28 ; Search time 1364.94 Seconds
                                    (without alignments)
                                    7703.105 Million cell updates/sec

Title:           US-09-939-408A-20
Perfect score:   2475
Sequence:        1 atggccccctccntcgactc..........nnnnnnnnnnnnnnnnnnn 2475

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:      6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_29Jan04:*
                 1:  geneseqn1980s:*
                 2:  geneseqn1990s:*
                 3:  geneseqn2000s:*
                 4:  geneseqn2001as:*
                 5:  geneseqn2001bs:*
                 6:  geneseqn2002s:*
                 7:  geneseqn2003as:*
                 8:  geneseqn2003bs:*
                 9:  geneseqn2003cs:*
                 10:  geneseqn2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2128	86.0	2475	6	ABA95244	Aba95244 Rhodotoru
2	2128	86.0	2475	6	AAD45816	Aad45816 Phenylala
3	1944	78.5	2163	6	AAD45824	Aad45824 Yeast phe
4	1699.4	68.7	2439	6	ABA95243	Aba95243 R. torulo
5	1699.4	68.7	2439	6	AAD45815	Aad45815 Rhodotoru
6	1677.4	67.8	2331	1	AAN81101	Aan81101 Sequence
7	1677.4	67.8	2331	1	AAN81116	Aan81116 L-phenyla
8	1675.8	67.7	2328	1	AAN81399	Aan81399 Sequence
9	1629	65.8	2419	6	ABA95240	Aba95240 R. gramin
10	1629	65.8	2419	6	AAD45812	Aad45812 Rhodotoru
11	1626.4	65.7	2151	6	AAD33076	Aad33076 Rhodotoru
12	1626.4	65.7	2151	6	AAD33100	Aad33100 R. glutin
13	1626.4	65.7	2151	6	AAD26924	Aad26924 Rhodotoru
14	1626.4	65.7	2151	7	ABS56543	Abs56543 R. glutin
15	1626.4	65.7	2151	9	ADE36055	Ade36055 Rhodospor
16	1624.8	65.6	2151	6	AAD33152	Aad33152 R. glutin
17	1624.8	65.6	2151	6	AAD33097	Aad33097 R. glutin
18	1624.8	65.6	2151	9	ADE36061	Ade36061 Rhodospor
19	1624.8	65.6	2151	9	ADE36075	Ade36075 Rhodospor
20	1624.8	65.6	2151	9	ADE36065	Ade36065 Rhodospor
21	1623.2	65.6	2151	6	AAD33099	Aad33099 R. glutin
22	1623.2	65.6	2151	6	AAD33096	Aad33096 R. glutin
23	1623.2	65.6	2151	6	AAD33075	Aad33075 Rhodotoru
24	1623.2	65.6	2151	6	AAD33098	Aad33098 R. glutin
25	1623.2	65.6	2151	6	AAD26923	Aad26923 Rhodotoru
26	1623.2	65.6	2151	7	ABS56530	Abs56530 R. glutin
27	1623.2	65.6	2151	9	ADE36053	Ade36053 Rhodospor
28	1623.2	65.6	2151	9	ADE36059	Ade36059 Rhodospor
29	1623.2	65.6	2151	9	ADE36067	Ade36067 Rhodospor
30	1623.2	65.6	2151	9	ADE36073	Ade36073 Rhodospor
31	1623.2	65.6	2151	9	ADE36057	Ade36057 Rhodospor
32	1621.6	65.5	2151	9	ADE36063	Ade36063 Rhodospor
33	1621.6	65.5	2151	9	ADE36069	Ade36069 Rhodospor
34	1620	65.5	2151	1	AAN91092	Aan91092 Structura
35	1618.4	65.4	2151	9	ADE36071	Ade36071 Rhodospor
36	1610.4	65.1	2061	6	AAD33095	Aad33095 Rhodotoru
37	1598.6	64.6	2787	6	ABA95242	Aba95242 R. rubra/
38	1598.6	64.6	2787	6	AAD45814	Aad45814 Rhodotoru
39	1584.8	64.0	2085	1	AAN82309	Aan82309 Intron-fr
40	1194.6	48.3	2987	1	AAN80456	Aan80456 Sequence
41	1124.4	45.4	2741	6	ABA95251	Aba95251 R. gramin
42	1124.4	45.4	2741	6	AAD45823	Aad45823 Rhodotoru
43	888.8	35.9	1355	1	AAN80457	Aan80457 Sequence
44	852.4	34.4	1218	1	AAN80458	Aan80458 Sequence
45	447	18.1	2280	7	ADA70967	Ada70967 Rice gene

ALIGNMENTS

```
RESULT 1
ABA95244
ID   ABA95244 standard; DNA; 2475 BP.
XX
AC   ABA95244;
XX
DT   29-AUG-2003  (revised)
DT   07-AUG-2003  (revised)
DT   10-JUN-2002  (first entry)
XX
DE   Rhodotorula PAL consensus polynucleotide sequence.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
KW   consensus; ds.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..2475
FT                   /*tag=  a
FT                   /note= "n= A or C or G or T; "n" indicates no consensus
FT                   at that position"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   P-PSDB; ABB07694.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
```

```
XX
PS   Claim 5; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a consensus polynucleotide sequence of the Rhodotorula PAL.
CC   (Updated on 07-AUG-2003 to correct OS field.) (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;

  Query Match             86.0%;  Score 2128;  DB 6;  Length 2475;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-288;
  Matches 2408;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
```

```
Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy       2401 NNNANAAA 2408
              ||||||||
Db       2401 NNNANAAA 2408
```

RESULT 2
AAD45816

```
ID   AAD45816 standard; DNA; 2475 BP.
XX
AC   AAD45816;
XX
DT   07-AUG-2003  (revised)
DT   27-DEC-2002  (first entry)
XX
DE   Phenylalanine ammonia lyase consensus DNA.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; ds.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 5; Page 45-46; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is PAL consensus DNA.
CC   (Updated on 07-AUG-2003 to correct OS field.)
XX
SQ   Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;

  Query Match             86.0%;  Score 2128;  DB 6;  Length 2475;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-288;
  Matches 2408;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        301 AAGATCGACAAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGATCGACAAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
```

```
Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNNCNNNC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNNCNNNC 2340

Qy       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy       2401 NNNANAAA 2408
              ||||||||
Db       2401 NNNANAAA 2408
```

```
RESULT 3
AAD45824
ID   AAD45824 standard; DNA; 2163 BP.
XX
AC   AAD45824;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Yeast phenylalanine ammonia lyase DNA #2.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; ds.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2163
FT                   /*tag= a
FT                   /product= "PAL protein"
FT                   /EC_number= "4.3.1.5"
FT                   /transl_except= (pos:13..15, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:34..36, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:46..48, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:49..51, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
```

```
FT                   /transl_except= (pos:55. .57, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:58. .60, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser, Leu or an in-frame stop
FT                   codon"
FT                   /transl_except= (pos:73. .75, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Pro, Ser"
FT                   /transl_except= (pos:79. .81, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro, Ser"
FT                   /transl_except= (pos:82. .84, aa:Xaa)
FT                   /note= "Xaa corresponds to Ala, Pro"
FT                   /transl_except= (pos:100. .102, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr"
FT                   /transl_except= (pos:106. .108, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Ser"
FT                   /transl_except= (pos:115. .117, aa:Xaa)
FT                   /note= "Xaa corresponds to Ala, Pro, Ser"
FT                   /transl_except= (pos:118. .120, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Gly, Trp"
FT                   /transl_except= (pos:142. .144, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Thr, Met, Glu, Ala, Val,
FT                   Gln, Pro, Leu"
FT                   /transl_except= (pos:160. .162, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:166. .168, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Glu, Gln"
FT                   /transl_except= (pos:193. .195, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Asp, Val"
FT                   /transl_except= (pos:196. .198, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Val, Leu"
FT                   /transl_except= (pos:226. .228, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:259. .261, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Pro, Ser"
FT                   /transl_except= (pos:277. .279, aa:Xaa)
FT                   /note= "Xaa corresponds to Asp, Ala"
FT                   /transl_except= (pos:304. .306, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Asn"
FT                   /transl_except= (pos:307. .309, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Ser"
FT                   /transl_except= (pos:325. .327, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:334. .336, aa:Xaa)
FT                   /note= "Xaa corresponds to Asp, His, Tyr"
FT                   /transl_except= (pos:340. .342, aa:Xaa)
FT                   /note= "Xaa corresponds to Arg, Ser"
FT                   /transl_except= (pos:349. .351, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:442. .444, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:448. .450, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:457. .459, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Val, Phe"
FT                   /transl_except= (pos:460. .462, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Asp"
FT                   /transl_except= (pos:469. .471, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser, Gly, Arg"
FT                   /transl_except= (pos:475. .477, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:547. .549, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr"
FT                   /transl_except= (pos:667. .669, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:673. .675, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser"
FT                   /transl_except= (pos:709. .711, aa:Xaa)
FT                   /note= "Xaa corresponds to Ser, Thr, Ile, Gly, Ala, Val,
FT                   Cys, Phe"
FT                   /transl_except= (pos:715. .717, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:721. .723, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:724. .726, aa:Xaa)
FT                   /note= "Xaa corresponds to Val, Leu, Phe"
FT                   /transl_except= (pos:736. .738, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Asn, Thr"
FT                   /transl_except= (pos:751. .753, aa:Xaa)
FT                   /note= "Xaa corresponds to Tyr, Ser, Phe"
FT                   /transl_except= (pos:775. .777, aa:Xaa)
FT                   /note= "Xaa corresponds to Leu, Phe"
FT                   /transl_except= (pos:913. .915, aa:Xaa)
FT                   /note= "Xaa corresponds to Leu"
FT                   /transl_except= (pos:955. .957, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1036. .1038, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Arg, Thr, Gln, Pro, Trp,
FT                   Ser or a stop codon"
FT                   /transl_except= (pos:1231. .1233, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Thr, Met"
FT                   /transl_except= (pos:1261. .1263, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1369. .1371, aa:Xaa)
FT                   /note= "Xaa corresponds to Leu"
FT                   /transl_except= (pos:1372. .1374, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1396. .1398, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1459. .1461, aa:Xaa)
FT                   /note= "Xaa corresponds to Gly"
FT                   /transl_except= (pos:1477. .1479, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:1498. .1500, aa:Arg)
FT                   /transl_except= (pos:1552. .1554, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr"
FT                   /transl_except= (pos:1606. .1608, aa:Xaa)
FT                   /note= "Xaa corresponds to Val"
FT                   /transl_except= (pos:1666. .1668, aa:Xaa)
FT                   /note= "Xaa corresponds to Met, Ile, Val, Leu"
FT                   /transl_except= (pos:1669. .1671, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Val, Leu"
FT                   /transl_except= (pos:1672. .1674, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Pro"
FT                   /transl_except= (pos:1675. .1677, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1681. .1683, aa:Xaa)
FT                   /note= "Xaa corresponds to Ile, Leu, Phe"
FT                   /transl_except= (pos:1684. .1686, aa:Xaa)
FT                   /note= "Xaa corresponds to Lys, Glu, Gln"
FT                   /transl_except= (pos:1699. .1701, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1714. .1716, aa:Xaa)
FT                   /note= "Xaa corresponds to Asn, Tyr"
FT                   /transl_except= (pos:1723. .1725, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Ala, Val"
FT                   /transl_except= (pos:1741. .1743, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:1747. .1749, aa:Xaa)
FT                   /note= "Xaa corresponds to Asn, Asp, Tyr"
FT                   /transl_except= (pos:1849. .1851, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro"
FT                   /transl_except= (pos:1852. .1854, aa:Xaa)
FT                   /note= "Xaa corresponds to Pro, Ser"
FT                   /transl_except= (pos:1927. .1929, aa:Xaa)
FT                   /note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
FT                   stop codon"
FT                   /transl_except= (pos:1939. .1941, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Pro, Ser"
FT                   /transl_except= (pos:1948. .1950, aa:Xaa)
FT                   /note= "Xaa corresponds to Ala, Pro, Ser"
FT                   /transl_except= (pos:1978. .1980, aa:Xaa)
FT                   /note= "Xaa corresponds to Thr, Ala, Ser"
FT                   /transl_except= (pos:2122. .2124, aa:Xaa)
FT                   /note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
FT                   Tyr, Cys"
FT                   /transl_except= (pos:2128. .2130, aa:Xaa)
```

```
FT                   /note= "Xaa corresponds to Ser, Arg, Cys"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   P-PSDB; AAE27943.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 7; Page 53-56; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.

  Query Match             78.5%;  Score 1944;  DB 6;  Length 2163;
  Best Local Similarity   93.2%;  Pred. No. 1.2e-262;
  Matches 2037;  Conservative    0;  Mismatches  126;  Indels   23;  Gaps    7;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCCCCCTCCBTCGACTCGATCGCGACCTCGBTCGCCAACGGCBTCDCVAACGGVTHG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CACGCCGCTCCGHCCAAGYCGSCWACGGGCGCCACGTCCACHCTCMGMCGGCCGBC-CDG 119

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        120 GCTCGCTCCTCCCG-CCACCCAGVHGACGCAGCTCGACATCGTBGAGVAGATCCTCGCCG 178

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        179 ACCCCACCGAC---GACGWSVTCGAACTCGACGGGTACACCCTCACCCTCGGHGACGTCG 235

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        236 TCGGCGCCGCBCGCAAGGGCCGCHCBGTCCGCGTC-CAGACAGMCGACGAGATCCGCGCA 294

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        295 AAGATCGACAAVAGBGTCGAGTTCCTCCGBDCBCAGCTCBACAACAGBGTCTACGGHGTC 354

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        355 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCVATCTCGCTCCAGAAG 414

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        415 GCBCTCCTCGAGCACCAGCTCTGCGGTGTBCTCCCBACGTCGDTCGABTCCTTCVGCCTC 474

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        475 GGHCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 534

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        535 GTCAACTCGCTCACBCGCGGCCACTCGGCBGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 594

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        595 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 654

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        655 GGCGACCTCTCCCCBCTCTCBTACATCGCCGCCGCCATCACCGGTCACCCGGACDBCAAG 714

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        715 GTBCACGTYKTSCACGAGGGCAMSGAGAAGATCATGTHCGCCCGCGAGGCGATCGCGCTC 774

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        775 TTBGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 834

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        835 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 894

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        895 TCGCAGGCGCTCACGGCTCTBACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 954

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        955 CCVTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1014

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db       1015 CGCACGCTCCTCGAGGGCAGCHVGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1074

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1075 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1134

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1135 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCBGTCCTCTCGCTCGAGGCCG---AG 1191

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1192 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGAHGACCCACCACGGCGGCAAC 1251

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1252 TTCCAGGCGDCCGCTGTCGCVAACACGATGGAGAAGACTCGCCTCGCVCTCGCCCTGATC 1311

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1312 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTB 1371

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1372 CCBTCCTGCCTCGCTGCCGAGGACCCBTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1431

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1432 GCCGCBGCBGCBTACACTTCGGAGCTCGGHCACCTTGCCAACCCGGTBACGACCCACGTC 1491

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1492 CAGCCGGCHGAGATGGGCAACCAGGCCGTCAACTCGCTCGCBCTCATCTCGGCBCGCCGC 1551

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1552 ACBGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTBCTC 1611

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1612 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGVTBVTC 1671

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
```

```
Db        1672 VCBDCGCTCHTCVAGCAGCACTTTGGCDCYGCCCTCGACGGC-----------WACGAA 1719

Qy        1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Db        1720 CTCGHGGACAAGGTCAACAAGDCGCTCDACAAGCGBCTCGAGCAGACCAACTCGTACGAC 1779

Qy        1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Db        1780 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGCTCCTC 1839

Qy        1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Db        1840 TCGTCCTCGCCBYCTGCCA--AGGTCTCGCTCGCBGCCGTCAACGCCTGGAAGGTCGCCT 1897

Qy        1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db        1898 CCGCCGAGAAGGCCATCTCGCTCACGCGCBAVGTCCGCGACHCCTTCTGGBCGGCBCCGT 1957

Qy        1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db        1958 CGTCGTCGTCGCCCGCGCTCDCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2017

Qy        2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db        2018 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2077

Qy        2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db        2078 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGDVCGGCHGCATCAACC 2137

Qy        2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186

Db        2138 ACGTCCTCGTCAAGATGCTCGCDTAG 2163


RESULT 4
ABA95243
ID   ABA95243 standard; cDNA; 2439 BP.
XX
AC   ABA95243;
XX
DT   07-AUG-2003  (revised)
DT   10-JUN-2002  (first entry)
XX
DE   R. toruloides PAL polypeptide encoding cDNA.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate; ss;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "phenylalanine ammonia lyase"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   P-PSDB; ABB07693.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Example 3; Fig 2 ; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a R. toruloides PAL polypeptide encoding cDNA. (Updated on 07-
CC   AUG-2003 to correct OS field.)
XX
SQ   Sequence 2439 BP; 440 A; 920 C; 631 G; 448 T; 0 U; 0 Other;

  Query Match             68.7%;  Score 1699.4;  DB 6;  Length 2439;
  Best Local Similarity   79.4%;  Pred. No. 1.4e-228;
  Matches 1913;  Conservative    0;  Mismatches  459;  Indels   36;  Gaps    6;

Qy           1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Db           1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy          61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Db          61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy         121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Db         101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy         181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Db         161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy         241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Db         218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy         301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Db         277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy         361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Db         337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy         421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Db         397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy         481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Db         457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy         541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Db         517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy         601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Db         577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy         661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db         637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy         721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
```

```
Db      697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db      757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db      817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy      901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Db      877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Db      937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Db      997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy     1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Db     1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy     1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Db     1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy     1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Db     1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy     1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Db     1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy     1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Db     1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy     1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Db     1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Db     1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Db     1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Db     1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Db     1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Db     1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Db     1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Db     1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Db     1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db     1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db     1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db     2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db     2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Db     2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy     2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Db     2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy     2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNNCNNNC 2340

Db     2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCCGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy     2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Db     2305 AGTACCCATCCACGCATCCGATAAATCAGGGAGAGAATCTACGCTTGCGGGAGCTTCTTG 2364

Qy     2401 NNNANAAA 2408

Db     2365 CGCATAAA 2372
```

RESULT 5
AAD45815
ID   AAD45815 standard; DNA; 2439 BP.
XX
AC   AAD45815;
XX
DT   07-AUG-2003 (revised)
DT   27-DEC-2002 (first entry)
XX
DE   Rhodotorula toruloides PAL DNA.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; gene; ds.
XX
OS   Rhodosporidium toruloides.
XX
FH   Key   Location/Qualifiers
FT   CDS   1. .2151
FT   /*tag= a
FT   /product= "PAL protein"
FT   /EC_number= "4.3.1.5"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.

```
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   P-PSDB; AAE27941.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Example 3; Page 41-43; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is R. toruloides PAL DNA.
CC   (Updated on 07-AUG-2003 to correct OS field.)
XX
SQ   Sequence 2439 BP; 440 A; 920 C; 631 G; 448 T; 0 U; 0 Other;

  Query Match             68.7%;  Score 1699.4;  DB 6;  Length 2439;
  Best Local Similarity   79.4%;  Pred. No. 1.4e-228;
  Matches 1913;  Conservative    0;  Mismatches  459;  Indels   36;  Gaps    6;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716
```

```
Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db       2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db       2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCCGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy       2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db       2305 AGTACCCATCCACGCATCCGATAAATCAGGGAGAGAATCTACGCTTGCGGGAGCTTCTTG 2364

Qy       2401 NNNANAAA 2408
Db       2365 CGCATAAA 2372
```

RESULT 6
AAN81101
ID   AAN81101 standard; DNA; 2331 BP.
XX
AC   AAN81101;
XX
DT   27-AUG-2003 (revised)
DT   25-MAR-2003 (revised)
DT   19-NOV-1990 (first entry)
XX
DE   Sequence encoding L-phenylalanine ammonia lyase (PAL) from Rhodosporidium
DE   turoloides.
XX
KW   Cinnamic acid; expression vector; MT 10414 FerumP-8876; pYtrp6; ss.
XX
OS   Rhodosporidium toruloides.
XX
PN   EP278706-A.
XX
PD   17-AUG-1988.
XX
PF   08-FEB-1988;   88EP-00301011.
XX
PR   06-FEB-1987;   87JP-00024705.
PR   18-JUN-1987;   87JP-00152357.
XX
PA   (MITK) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N, Yoshino S, Sone S, Nakajima Y, Makiguchi N;
XX
DR   WPI; 1988-229543/33.
XX
PT   Recombinant plasmid for expression of L-phenylalanine ammonia-lyase -
PT   having combined promoter of tac promoter and Pl promoter for enhanced
PT   expression.
XX
PS   Example; Fig 2(A)-2(D); 37pp; English.
XX
CC   Plasmids were from E.coli MT 10414 (FermP-8876) carrying the plasmid
CC   pYtrp6 contg. the structural gene for PAL from Rhodosporidium toruloides.
CC   Inventors claim a recombinant plasmid contg. a DNA sequence coding for
CC   PAL with the SQ in AAP80501. Procedure for cloning the structural gene
CC   for PAL (AAN81101) is described as a reference example. The hybrid
CC   plasmid permits more efficient expression of PAL in E.coli. PAL is used
CC   for the prodn. of L-phenylalanine from cinnamic acid and ammonia (Updated
CC   on 25-MAR-2003 to correct PR field.) (Updated on 27-AUG-2003 to correct
CC   OS field.)
XX
SQ   Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;

Query Match 67.8%; Score 1677.4; DB 1; Length 2331;
Best Local Similarity 80.4%; Pred. No. 1.6e-225;
Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6;

```
Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
```

```
Db      577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy      661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db      637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy      721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db      697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db      757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db      817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy      901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db      877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db      937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db      997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy     1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db     1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy     1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db     1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy     1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db     1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy     1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db     1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy     1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db     1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy     1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db     1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db     1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db     1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db     1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db     1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db     1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db     1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db     1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db     1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db     1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db     1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db     2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db     2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db     2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy     2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db     2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy     2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db     2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy     2341 NGTNNNCANNNA 2352
Db     2305 AGTACCCATCCA 2316
```

[redacted]

238/

RESULT 7
AAN81116
ID   AAN81116 standard; DNA; 2331 BP.
XX
AC   AAN81116;
XX
DT   24-OCT-2003 (revised)
DT   25-MAR-2003 (revised)
DT   12-NOV-1990 (first entry)
XX
DE   L-phenylalanine ammonialyase structural gene.
XX
KW   L-phenylalanine ammonialyase; PAL; foreign gene expression;
KW   culture temperature; expression regulation; ss.
XX
OS   Rhodosporidium toruloides; (IFO 559).
XX
FH   Key   Location/Qualifiers
FT   CDS   1. .2151
FT   /*tag= a
FT   /product= "PAL"
XX
PN   EP279665-A.
XX
PD   24-AUG-1988.
XX
PF   18-FEB-1988;   88EP-00301356.

106

```
XX
PR   19-FEB-1987;   87JP-00034397.
PR   18-JUN-1987;   87JP-00152359.
XX
PA   (MITK ) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N,  Yoshino S,  Yamamoto K,  Sone S,  Suzuki M,  Nakajima Y;
XX
DR   WPI; 1988-236895/34.
DR   P-PSDB; AAP80513.
XX
PT   Regulation of expression of foreign gene in Escherichia coli - by
PT   maintaining temp. of culture at 40 deg. C or more to suppress expression.
XX
PS   Disclosure; ); 36pp; English.
XX
CC   The cloned DNA from pSW13 isolated from the cloned cell carrying pSW13
CC   was digested with various restriction endonucleases. The sequences from
CC   the fragments were analysed by the method of Maxam-Gilbert and also by
CC   biochemical means using the method of Maat. The sequences obtained were
CC   edited using the GENAS program produced by Mitsui Information Development
CC   Co. The sequence so determined is represented here. The regulation of
CC   expression of this foreign gene in E.coli carrying a recombinant plasmid
CC   comprises maintaining the culture at at least 40 deg. C. See also EP-
CC   279664. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
CC   2003 to correct PI field.) (Updated on 24-OCT-2003 to standardise OS
CC   field)
XX
SQ   Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;

  Query Match             67.8%;  Score 1677.4;  DB 1;  Length 2331;
  Best Local Similarity   80.4%;  Pred. No. 1.6e-225;
  Matches 1891;  Conservative    0;  Mismatches  425;  Indels   36;  Gaps    6;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656
```

```
Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db       2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db       2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy       2341 NGTNNNCANNNA 2352
Db       2305 AGTACCCATCCA 2316
```

RESULT 8
AAN81399

```
ID   AAN81399 standard; cDNA; 2328 BP.
XX
AC   AAN81399;
XX
DT   12-DEC-1990  (first entry)
XX
DE   Sequence of Rhodosporidium toruloides IFO 559 cells cDNA encoding L-
DE   Phenylalanine ammonia lyase (PAL).
XX
KW   Aspartame; artificial sweetener; enzyme; ss.
XX
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT   misc_feature    1. .2148
FT                   /*tag=  b
FT                   /note= "Claimed"
XX
PN   EP260919-A.
XX
PD   23-MAR-1988.
XX
PF   15-SEP-1987;   87EP-00308119.
XX
PR   16-SEP-1986;   86JP-00215864.
XX
PA   (MITK ) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N,  Yoshino S,  Yamamoto K,  Se T,  Sone S,  Nakajima Y;
PI   Suzuki M,  Makiguchi N;
XX
DR   WPI; 1988-079223/12.
DR   P-PSDB; AAP81099.
XX
PT   New L-phenylalanine ammonia lyase and encoding DNA sequence - useful for
PT   producing L-phenylalanine from cinnamic acid and ammonia.
XX
PS   Disclosure; 9; 37pp; English.
XX
CC   PAL is useful in prodn. of L-phenylalanine from cinnamic acid and an NH3
CC   donor (process claimed). L-Phe is an essential AA and starting material
CC   for sweetener aspartame. E.coli transformants will produce PAL without
CC   induction by expensive AAs, no surfactant is needed to increase cell wall
CC   permeability
XX
SQ   Sequence 2328 BP; 423 A; 889 C; 591 G; 425 T; 0 U; 0 Other;

  Query Match            67.7%;  Score 1675.8;  DB 1;  Length 2328;
  Best Local Similarity  80.4%;  Pred. No. 2.7e-225;
  Matches 1890;  Conservative    0;  Mismatches  426;  Indels   36;  Gaps    6;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
```

```
Db     577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy     661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db     637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy     721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db     697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy     781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db     757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy     841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db     817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy     901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Db     877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy     961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Db     937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy    1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Db     997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGCAGGTCAAGGTCAAG 1056

Qy    1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Db    1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy    1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Db    1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy    1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Db    1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy    1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Db    1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy    1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Db    1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy    1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Db    1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy    1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Db    1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy    1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Db    1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy    1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Db    1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy    1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Db    1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy    1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Db    1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy    1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Db    1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy    1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Db    1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy    1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Db    1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy    1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db    1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy    1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db    1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy    2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db    2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy    2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db    2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy    2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Db    2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy    2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Db    2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy    2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Db    2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCGGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy    2341 NGTNNNCANNNA 2352

Db    2305 AGTACCCATCCA 2316
```

```
RESULT 9
ABA95240
ID   ABA95240 standard; cDNA; 2419 BP.
XX
AC   ABA95240;
XX
DT   10-JUN-2002  (first entry)
XX
DE   R. graminis PAL polypeptide encoding cDNA.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate; ss;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             37. .2199
FT                   /*tag=  a
FT                   /transl_except= (pos: 493. .495, aa: Xaa)
FT                   /note= "Xaa = Val or Ala"
FT                   /product= "phenylalanine ammonia lyase"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
```

```
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   P-PSDB; ABB07690.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Claim 4; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a R. graminis PAL polypeptide encoding cDNA
XX
SQ   Sequence 2419 BP; 416 A; 909 C; 659 G; 434 T; 0 U; 1 Other;

  Query Match             65.8%;  Score 1629;  DB 6;  Length 2419;
  Best Local Similarity   78.0%;  Pred. No. 9.1e-219;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db       1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707
```

```
Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC------------AACGCA 1755

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db       2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db       2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db       2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352

Qy       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db       2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411

Qy       2401 NNNANAAA 2408
Db       2412 AAAAAAAA 2419
```

```
RESULT 10
AAD45812
ID   AAD45812 standard; DNA; 2419 BP.
XX
AC   AAD45812;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Rhodotorula graminis PAL DNA #1.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide; gene; ds.
XX
OS   Rhodotorula graminis.
XX
FH   Key             Location/Qualifiers
FT   CDS             37. .2199
FT                   /*tag= a
FT                   /product= "PAL protein"
FT                   /EC_number= "4.3.1.5"
FT                   /transl_except= (pos:493. .495, aa:Xaa)
FT                   /note= "Xaa corresponds to Val or Ala"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   P-PSDB; AAE27938.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 13; Page 27-29; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is R. graminis PAL DNA
XX
SQ   Sequence 2419 BP; 416 A; 909 C; 659 G; 434 T; 0 U; 1 Other;

  Query Match             65.8%;  Score 1629;  DB 6;  Length 2419;
  Best Local Similarity   78.0%;  Pred. No. 9.1e-219;
  Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510
```

```
Qy      481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db      511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570

Qy      541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db      571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630

Qy      601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db      631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690

Qy      661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db      691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750

Qy      721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db      751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810

Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db      811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db      871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930

Qy      901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db      931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990

Qy      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db      991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050

Qy     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db     1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110

Qy     1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db     1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170

Qy     1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db     1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227

Qy     1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db     1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287

Qy     1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db     1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347

Qy     1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db     1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407

Qy     1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db     1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467

Qy     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db     1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527

Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db     1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db     1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db     1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740
Db     1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC-----------AACGCA 1755

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db     1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db     1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db     1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db     1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993

Qy     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db     1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db     2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db     2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db     2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232

Qy     2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db     2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292

Qy     2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db     2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352

Qy     2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db     2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411

Qy     2401 NNNANAAA 2408
Db     2412 AAAAAAAA 2419
```

RESULT 11
AAD33076

```
ID   AAD33076 standard; DNA; 2151 BP.
XX
AC   AAD33076;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant DNA.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
```

```
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; gene; mutant; variant; ds.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "Rhodotorula glutinis mutant PAL"
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   P-PSDB; AAE20654.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Example 10; Page 95-98; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL mutant DNA
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 2.1e-218;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296
```

```
Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

```
RESULT 12
AAD33100
ID   AAD33100 standard; DNA; 2151 BP.
XX
AC   AAD33100;
XX
DT   01-JUL-2002  (first entry)
XX
DE   R. glutinis phenylalanine ammonia lyase (PAL) mutant DNA, EP18Km-6.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; mutant; variant; gene; ds.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "Rhodotorula glutinis PAL mutant"
FT                   /EC_number= "4.3.1.5"
FT   mutation        replace(645, G)
FT                   /*tag=  b
FT   mutation        replace(792, A)
FT                   /*tag=  c
FT   mutation        replace(858, T)
FT                   /*tag=  d
FT   mutation        replace(1619, T)
FT                   /*tag=  e
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   P-PSDB; AAE20668.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Disclosure; Page; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL mutant DNA. Note: This sequence is not shown in the specification,
CC   however it is constructed based on the PAL DNA shown as SEQ.ID.NO:7
CC   (AAD33075) in the sequence listing
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 2.1e-218;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
```

```
Db      1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy     61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db     61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy    121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db    101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy    181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db    161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy    241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db    218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy    301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db    277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy    361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db    337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy    421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db    397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy    481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db    457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy    541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db    517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy    601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db    577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy    661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db    637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy    721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db    697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy    781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db    757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy    841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db    817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db    997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

```
RESULT 13
AAD26924
ID   AAD26924 standard; DNA; 2151 BP.
XX
AC   AAD26924;
XX
DT   09-APR-2002  (first entry)
XX
DE   Rhodotorula glutinis mutant phenylalanine ammonia-lyase (PAL) DNA.
XX
KW   Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
KW   para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW   Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW   aerospace application; chemical; red yeast; mutant; ds.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "R. glutinis mutant PAL enzyme"
FT   mutation        replace(1619, T)
FT                   /*tag=  b
XX
PN   WO200111071-A2.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US021156.
XX
PR   06-AUG-1999;  99US-0147719P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Tang X,  Vannelli TM,  Qi WW,  Sariaslani S,  Gatenby AA;
XX
DR   WPI; 2002-121549/16.
DR   P-PSDB; AAE16390.
XX
PT   Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT   comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia
PT   -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT   activity.
XX
PS   Claim 13; Page 67-68; 75pp; English.
XX
CC   The invention relates to a method for producing para-hydroxycinnamic acid
CC   (PHCA). The method comprises converting cinnamate to PHCA by converting
CC   glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC   route or by generating a new biocatalyst possessing enhanced tyrosine
CC   ammonia-lyase (TAL) activity. The method is useful for the biological
CC   production of PHCA which is useful as a monomer for the production of
CC   Liquid Crystal Polymers (LCP), where the LCP is used in electronic
CC   connectors, telecommunications and aerospace applications. LCP is also
CC   useful in medical devices, as well as chemical, and food packing
CC   applications due to its resistance to sterilising radiation. The present
CC   sequence is a DNA encoding red yeast, Rhodotorula glutinis
CC   (Rhodosporidium toruloides) mutant PAL derived by substituting Ile at
CC   position 540 with Thr. Mutant PAL enzyme posseses enhanced TAL activity
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             65.7%;  Score 1626.4;  DB 6;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 2.1e-218;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
```

```
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
              |||||||||||||||| |||||||||||||||||||||||    | |||||||||||||
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
              |||||||  ||||||| || ||||| |||||||||||||||||  ||||||| ||||
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
              || |||||||||| ||||  |||||  |||||||||| |||||||||||||||||||||
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
              ||||| || || ||||| |||||| | || ||||| |||||||| || ||||| || |||
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
              |||||||| |||||||  ||||||| |||||||||| || |||||||||| || ||| |||
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
              ||  ||||| ||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
              || ||| ||||| | ||||||  |||||||||||||||||||||||| |||       ||
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
                   ||||| || | |||||||||||| | ||| |    ||||               |||
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
              ||||  || ||||| |||||| |||||  |||||| ||||||||||||||||||||||
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
              ||||  |||||||||||||||||||| |||||  |||||||||||||||||
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
                 ||| || |                 | ||||||||| ||||||||||||||||||||
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
              |||||||  ||||||||||||| ||| | ||||||||  |||||||| | || |||
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
              || | |||||||||||||| ||||||||||||||||| |    |||| ||| | ||||
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
              |||||||||||| |||||||||||||||||||  |||||||||||||||||||||| ||||
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
              |||||||||| | ||||||||   |||||||||||||||||    ||| | ||||||
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
              |||||||| |||||||||||| |||
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 14
ABS56543

```
ID   ABS56543 standard; DNA; 2151 BP.
XX
AC   ABS56543;
XX
DT   27-JAN-2003  (first entry)
XX
DE   R. glutinis DNA encoding PAL/TAL naturally occurring variant.
XX
KW   ds; gene; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase;
KW   PAH; para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
KW   tyrosine; cinnamic acid; PHCA; liquid crystal polymer.
XX
OS   Rhodtorula glutinis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .2151
FT                   /*tag=  a
FT                   /product= "PAL/TAL"
XX
PN   WO200290523-A2.
XX
PD   14-NOV-2002.
XX
PF   03-MAY-2002; 2002WO-US018551.
XX
PR   04-MAY-2001; 2001US-0288701P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Qi WW,  Sariaslani FS,  Tang X;
XX
DR   WPI; 2003-058928/05.
DR   P-PSDB; ABG71255.
XX
PT   Novel recombinant host for producing para-hydroxycinnamic acid and
PT   tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT   gene encoding phenylalanine hydroxylase activity.
XX
PS   Example 4; Page 66-67; 69pp; English.
XX
CC   The invention relates to a recombinant host comprising at least one gene
CC   encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
CC   encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
CC   (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC   providing the recombinant host; and (2) growing the recombinant organism
CC   in the presence of a fermentable carbon substrate and thus producing para
CC   -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC   providing a recombinant organism comprising at least one gene encoding a
CC   phenylalanine hydroxylase activity; and (2) growing the recombinant
CC   organism in the presence of a fermentable carbon substrate to produce
CC   tyrosine. The recombinant host cell is used for producing para-
CC   hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC   production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC   monomer for the production of liquid crystal polymer. The present
CC   sequence encodes a naturally occurring variant of the R. glutinis
CC   Phenylalanine ammonia-lyase, PAL, which has TAL activity (PAL/TAL)
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             65.7%;  Score 1626.4;  DB 7;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 2.1e-218;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
              ||||| ||||| |||||||||| ||  |||| |||| |||||| ||  |     ||   |
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              || || |                 |||||| || |||  |||          || | ||| |
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              |||||| ||| || || ||||||   |||||| ||||||||| ||| |||| |||||||
```

```
Db      101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy      181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db      161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy      241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db      218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy      301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db      277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy      361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db      337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy      421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db      397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy      481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db      457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy      541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db      517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy      601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db      577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy      661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db      637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy      721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db      697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db      757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db      817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy      901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db      877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db      937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db      997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy     1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db     1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy     1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db     1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy     1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db     1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy     1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db     1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy     1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db     1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy     1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db     1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db     1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db     1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db     1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db     1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db     1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db     1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db     1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db     1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db     1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy     1981 CGTCGTCGTCGCCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db     1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db     2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db     2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db     2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

```
RESULT 15
ADE36055
ID   ADE36055 standard; DNA; 2151 BP.
XX
AC   ADE36055;
XX
DT   29-JAN-2004  (first entry)
```

```
XX
DE   Rhodosporidium glutinis EP18Km-6 mutant PAL encoding DNA SEQ ID NO:3.
XX
KW   recombination; nucleic acid template; recombined polypeptide;
KW   altered property; combined gene; gene; ds.
XX
OS   Rhodosporidium.
XX
PN   WO2003072743-A2.
XX
PD   04-SEP-2003.
XX
PF   26-FEB-2003; 2003WO-US005708.
XX
PR   26-FEB-2002; 2002US-0360279P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Milano J,  Tang X;
XX
DR   WPI; 2003-721766/68.
DR   P-PSDB; ADE36056.
XX
PT   Recombination of nucleic acid templates, useful e.g. for producing
PT   enzymes with altered properties, is based on cyclic extension of unpaired
PT   primers.
XX
PS   Example 2; SEQ ID NO 3; 481pp; English.
XX
CC   The present invention describes a method for the recombination of nucleic
CC   acid (NA) templates. The method comprises: (a) providing at least two
CC   double-stranded templates (T1, T2) having different 5' and 3' regions in
CC   their sense strands; (b) contacting the templates with at least one each
CC   of forward and reverse primers that anneal, respectively, only to the 3'-
CC   region of the antisense strand of T1 and to the 3'-region of the sense
CC   strand of T2; (c) extending the primers by no more than 1000 nucleotides
CC   (nt); (d) separating extended primers from their templates; (e)
CC   reannealing the extended primers, to either template; and (f) repeating
CC   cycles of steps (b)-(e) until at least one full-length extension product,
CC   i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
CC   method starting from at least one each of antisense and sense single-
CC   stranded templates; (2) generating a recombined polypeptide (I) with
CC   altered properties by expressing recombination products of the new
CC   processes and screening the polypeptides formed for properties different
CC   from those of polypeptides from either of the templates; and (3) (I)
CC   produced by method (2). The methods are useful for the recombination of
CC   nucleic acid templates of interest, and generation of a recombined
CC   polypeptide having altered properties. The method can be used to make
CC   combined genes that express proteins e.g. enzymes, cytokines, growth
CC   factors, viral proteins or microbial antigens with altered properties,
CC   e.g. stability, activity or specificity. The present sequence is used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

  Query Match             65.7%;  Score 1626.4;  DB 9;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 2.1e-218;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296
```

```
Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

Search completed: September 11, 2004, 14:50:34
Job time : 1376.94 secs

```
OM nucleic - nucleic search, using sw model

Run on:         September 11, 2004, 13:38:08 ; Search time 185.6 Seconds
                                          (without alignments)
                                          7400.353 Million cell updates/sec

Title:          US-09-939-408A-20
Perfect score:  2475
Sequence:       1 atggcccccctccntcgactc..........nnnnnnnnnnnnnnnnnnn 2475

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:      1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2128	86.0	2475	4	US-09-624-693A-20	Sequence 20, Appl
	2	1699.4	68.7	2439	4	US-09-624-693A-18	Sequence 18, Appl
	3	1629.4	65.8	2419	4	US-09-624-693A-12	Sequence 12, Appl
	4	1626.4	65.7	2151	4	US-09-627-216A-9	Sequence 9, Appli
	5	1626.4	65.7	2151	4	US-09-765-873A-9	Sequence 9, Appli
	6	1623.2	65.6	2151	4	US-09-627-216A-7	Sequence 7, Appli
	7	1623.2	65.6	2151	4	US-09-765-873A-7	Sequence 7, Appli
	8	1610.4	65.1	2061	4	US-09-765-873A-31	Sequence 31, Appl
	9	1598.6	64.6	2787	4	US-09-624-693A-16	Sequence 16, Appl
	10	322.6	13.0	2311	4	US-09-624-693A-14	Sequence 14, Appl
	11	213.2	8.6	1455	4	US-09-615-192A-245	Sequence 245, App
	12	170.2	6.9	1520	4	US-09-615-192A-247	Sequence 247, App
	13	131.6	5.3	1548	4	US-09-252-991A-16475	Sequence 16475, A
c	14	131.6	5.3	2295	4	US-09-252-991A-16126	Sequence 16126, A
	15	111	4.5	577	4	US-09-615-192A-97	Sequence 97, Appl
	16	111	4.5	577	4	US-09-169-789-97	Sequence 97, Appl
	17	111	4.5	684	2	US-08-975-316-45	Sequence 45, Appl
	18	111	4.5	684	4	US-09-615-192A-45	Sequence 45, Appl
	19	111	4.5	684	4	US-09-169-789-45	Sequence 45, Appl
	20	107.4	4.3	1926	4	US-09-249-585A-4	Sequence 4, Appli
	21	107.4	4.3	1931	2	US-09-130-114-2	Sequence 2, Appli
	22	106.6	4.3	1527	4	US-09-489-039A-6404	Sequence 6404, Ap
	23	106.2	4.3	681	4	US-09-615-192A-244	Sequence 244, App
c	24	97.2	3.9	77536	4	US-09-410-551B-1	Sequence 1, Appli
	25	94.2	3.8	624	2	US-08-713-000-9	Sequence 9, Appli
	26	94.2	3.8	624	2	US-08-975-316-9	Sequence 9, Appli
	27	94.2	3.8	624	3	US-09-211-710-9	Sequence 9, Appli
	28	94.2	3.8	624	4	US-09-615-192A-9	Sequence 9, Appli
	29	94.2	3.8	624	4	US-09-169-789-9	Sequence 9, Appli
	30	92.2	3.7	28958	1	US-08-258-261B-6	Sequence 6, Appli
	31	92.2	3.7	28958	1	US-08-456-837-6	Sequence 6, Appli
	32	92.2	3.7	28958	1	US-08-457-342-6	Sequence 6, Appli
	33	92.2	3.7	28958	1	US-08-457-646A-6	Sequence 6, Appli
	34	92.2	3.7	28958	1	US-08-458-076A-6	Sequence 6, Appli
	35	92.2	3.7	28958	1	US-08-764-233A-4	Sequence 4, Appli
	36	92.2	3.7	28958	1	US-08-457-335A-6	Sequence 6, Appli
	37	92.2	3.7	28958	1	US-08-729-214-6	Sequence 6, Appli
	38	92.2	3.7	28958	3	US-09-028-934-6	Sequence 6, Appli
	39	92.2	3.7	49377	1	US-08-764-233A-1	Sequence 1, Appli
	40	91.6	3.7	5392	2	US-08-403-852D-1	Sequence 1, Appli
	41	91.6	3.7	5392	3	US-08-510-646B-1	Sequence 1, Appli
	42	91.6	3.7	5392	3	US-09-231-818-1	Sequence 1, Appli
	43	91.6	3.7	5392	4	US-09-635-359B-1	Sequence 1, Appli
	44	91	3.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	45	91	3.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-624-693A-20
; Sequence 20, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSC
;  CURRENT APPLICATION NUMBER: US/09/624,693A
;  CURRENT FILING DATE:  2000-07-24
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 2475
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_difference
;   LOCATION: (13, 34, 46, 49, 51, 57, 59, 68, 69, 73, 75 - 77, 79, 82, 84,
;   OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus
;   OTHER INFORMATION: Sequence of SEQ ID NOs: 12, 16, and 18
; Patent No. 6355468
US-09-624-693A-20

  Query Match            86.0%;  Score 2128;  DB 4;  Length 2475;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2408;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNNCNNNC 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNNCNNNC 2340

Qy       2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy       2401 NNNANAAA 2408
              ||||||||
Db       2401 NNNANAAA 2408
```

```
RESULT 2
US-09-624-693A-18
; Sequence 18, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSC
;  CURRENT APPLICATION NUMBER: US/09/624,693A
;  CURRENT FILING DATE:  2000-07-24
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 2439
;   TYPE: DNA
;   ORGANISM: Rhodotorula toruloides
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2148)
US-09-624-693A-18

  Query Match              68.7%;  Score 1699.4;  DB 4;  Length 2439;
  Best Local Similarity    79.4%;  Pred. No. 3.2e-305;
  Matches 1913;  Conservative    0;  Mismatches  459;  Indels   36;  Gaps    6;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
```

```
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db       2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db       2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCCGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy       2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNN 2400
Db       2305 AGTACCCATCCACGCATCCGATAAATCAGGGAGAGAATCTACGCTTGCGGGAGCTTCTTG 2364

Qy       2401 NNNANAAA 2408
Db       2365 CGCATAAA 2372
```

RESULT 3
US-09-624-693A-12
; Sequence 12, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 2419
; TYPE: DNA
; ORGANISM: Rhodotorula graminis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)..(2196)
US-09-624-693A-12

Query Match 65.8%; Score 1629.4; DB 4; Length 2419;
Best Local Similarity 78.0%; Pred. No. 2.6e-292;
Matches 1879; Conservative 0; Mismatches 504; Indels 25; Gaps 8;

```
Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db         37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271

Qy        241 TCGGCGCCGCNCGCAAGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGTCTCGTCCTTCAGCGTC 510

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        991 CCGTTCATCCACGACGTCTGCCGCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
```

```
Db       1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC------------AACGCA 1755

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db       2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232

Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db       2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNNCNNNC 2340
Db       2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352

Qy       2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db       2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411

Qy       2401 NNNANAAA 2408
Db       2412 AAAAAAAA 2419
```

RESULT 4
US-09-627-216A-9
; Sequence 9, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: mutant from
;   OTHER INFORMATION: Rhodotorula glutinis
US-09-627-216A-9

```
  Query Match             65.7%;  Score 1626.4;  DB 4;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 9.3e-292;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276
```

```
Qy    301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db    277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy    361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db    337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy    421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db    397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy    481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db    457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy    541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db    517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy    601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db    577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy    661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db    637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy    721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db    697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy    781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db    757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy    841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db    817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db    997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 5
US-09-765-873A-9
; Sequence 9, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9

```
;  LENGTH: 2151
;  TYPE: DNA
;  ORGANISM: mutant from Rhodotorula glutinis
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(2151)
US-09-765-873A-9

  Query Match             65.7%;  Score 1626.4;  DB 4;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 9.3e-292;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
```

```
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 6
US-09-627-216A-7
; Sequence 7, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodotorula glutinis
US-09-627-216A-7

```
  Query Match             65.6%;  Score 1623.2;  DB 4;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 3.6e-291;
  Matches 1827;  Conservative    0;  Mismatches  324;  Indels   35;  Gaps    5;

Qy      1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db      1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy     61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db     61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy    121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db    101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy    181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db    161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy    241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db    218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy    301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db    277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy    361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db    337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy    421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db    397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy    481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db    457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy    541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db    517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy    601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db    577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy    661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db    637 GGCGACCTGTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy    721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db    697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy    781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db    757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy    841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db    817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTC 876

Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db    997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536
```

```
Qy      1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db      1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy      1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGTNTNNTC 1680
Db      1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy      1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db      1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy      1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db      1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy      1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db      1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy      1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db      1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy      1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db      1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy      1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db      1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy      2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db      2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy      2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db      2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy      2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db      2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 7
US-09-765-873A-7
; Sequence 7, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2151)
US-09-765-873A-7

```
  Query Match             65.6%;  Score 1623.2;  DB 4;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 3.6e-291;
  Matches 1827;  Conservative    0;  Mismatches  324;  Indels   35;  Gaps    5;

Qy         1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db         1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy        61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db        61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy       121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db       101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy       181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db       161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy       241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db       218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy       301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db       277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy       361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db       337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy       421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db       397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy       481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db       457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy       541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db       517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy       601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db       577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy       661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db       637 GGCGACCTGTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy       721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db       697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy       781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db       757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy       841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db       817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTC 876

Qy       901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db       877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy       961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db       937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy      1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db       997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056
```

```
Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 8
US-09-765-873A-31
; Sequence 31, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
;   LENGTH: 2061
;   TYPE: DNA
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: CDS]
;   LOCATION: (1)..(2061)
US-09-765-873A-31

```
  Query Match             65.1%;  Score 1610.4;  DB 4;  Length 2061;
  Best Local Similarity   85.0%;  Pred. No. 8.2e-289;
  Matches 1764;  Conservative    0;  Mismatches  297;  Indels   15;  Gaps    3;

Qy        111 GCCGNCGCNGGCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAG 170
Db          1 GCAGTCGCAGGCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAG 60

Qy        171 ATCCTCGCCGACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTC 230
Db         61 ATGCTCGCCGCGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTC 117

Qy        231 GGNGACGTCGTCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGA 290
Db        118 GGAGACGTCGTCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGA 176

Qy        291 GATCCGCGCAAAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGT 350
Db        177 GATCCGCTCAAAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGT 236

Qy        351 CTACGGNGTCACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTC 410
Db        237 CTACGGCGTCACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTC 296

Qy        411 GCTCCAGAAGGCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTC 470
Db        297 GCTCCAGAAGGCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTC 356

Qy        471 CTTCNGCCTCGGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCAT 530
Db        357 GTTCCGCCTCGGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCAT 416

Qy        531 GACCATCCGCGTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGA 590
Db        417 GACAATCCGCGTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGA 476

Qy        591 GGCGCTCACCAACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCAT 650
Db        477 GGCGCTCACCAACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCAT 536

Qy        651 CTCGGCGTCGGGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCC 710
Db        537 CTCTGCGTCGGGCGACCTGTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCC 596
```

```
Qy     711 GGACNNCAAGGTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGC 770
Db     597 GGACAGCAAGGTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGC 656

Qy     771 GATCGCGCTCTTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGT 830
Db     657 GATGGCGCTCTTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGT 716

Qy     831 CAACGGCACGGCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCT 890
Db     717 CAACGGCACCGCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCT 776

Qy     891 CTCGCTCCTCTCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGG 950
Db     777 CTCGCTCCTCTCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGG 836

Qy     951 CTCGTTCCACCCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGC 1010
Db     837 CTCGTTCCACCCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGC 896

Qy    1011 GCGCAACATCCGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGT 1070
Db     897 GGGAAACATCCGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGT 956

Qy    1071 CAAGGTCAAGGACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCC 1130
Db     957 CAAGGTCAAGGACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCC 1016

Qy    1131 TCAGTGGCTCGGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGA 1190
Db    1017 TCAGTGGCTCGGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGA 1076

Qy    1191 GGCCGGNCAGTCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCA 1250
Db    1077 GGCCGGCCAGTCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCA 1136

Qy    1251 CGGCGGCAACTTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCT 1310
Db    1137 CGGCGGCAATTTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCT 1196

Qy    1311 CGCCCTGATCGGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAA 1370
Db    1197 CGCCCAGATCGGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAA 1256

Qy    1371 CCGCGGCCTNCCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGG 1430
Db    1257 CCGCGGCCTCCCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGG 1316

Qy    1431 CCTCGACATTGCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNAC 1490
Db    1317 CCTCGACATCGCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGAC 1376

Qy    1491 GACCCACGTCCAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTC 1550
Db    1377 GACGCATGTCCAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTC 1436

Qy    1551 GGCNCGCCGCACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTA 1610
Db    1437 GGCTCGTCGCACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTA 1496

Qy    1611 CTGCGTNCTCCAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGA 1670
Db    1497 CTGCGTTCTCCAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGG 1556

Qy    1671 CCCGNTNNTCNCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNN 1730
Db    1557 CCCAGCCATCGTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCT 1616

Qy    1731 NNNNNACGAACTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAA 1790
Db    1617 GCGCGACGAGCTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAA 1676

Qy    1791 CTCGTACGACCTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGT 1850
Db    1677 CTCGTACGACCTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGT 1736

Qy    1851 CGAGNNNNNNNNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGG 1910
Db    1737 CGAGGTCCTCTCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGG 1785

Qy    1911 AAGGTCGCCTCCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGG 1970
Db    1786 AAGGTCGCCGCCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGG 1845

Qy    1971 NCGGCNCCGTCGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTG 2030
Db    1846 TCCGCCGCGTCGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTC 1905

Qy    2031 TACTCGTTCGTCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGC 2090
Db    1906 TACGCCTTCGTCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGC 1965

Qy    2091 AAGCAGGAGGTGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGC 2150
Db    1966 AAGCAAGAGGTGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGC 2025

Qy    2151 NGCATCAACCACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db    2026 AGGATCAACAACGTCCTCCTCAAGATGCTCGCTTAG 2061
```

RESULT 9
US-09-624-693A-16
; Sequence 16, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 2787
;   TYPE: DNA
;   ORGANISM: Rhodotorula mucilaginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (646)..(2784)
US-09-624-693A-16

Query Match 64.6%; Score 1598.6; DB 4; Length 2787;
Best Local Similarity 82.9%; Pred. No. 1.3e-286;
Matches 1812; Conservative 0; Mismatches 330; Indels 44; Gaps 5;

```
Qy       1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db     646 ATGGCCCCCTCCGTCGACTCGATCGCGACTTCGGTTGCCAACTCCCTCTCGAACGGGTTG 705

Qy      61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db     706 CACG-----CCGCCGCCGCCGCCAACGGTGGCGACGTCCACAAGAAGACGGCCGGTGCTG 760

Qy     121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db     761 GCTCCCTCCTCCCGACCACCGAGACGACCCAGCTCGACATCGTTGAGCGCATCTTGGCCG 820

Qy     181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db     821 ACGCCGGCGCGACGGACCAGATCAAACTCGATGGGTACACCCTCACGCTCGGCGACGTCG 880
```

```
Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        881 TCGGCGCTGCTCGCCGTGGCCGCTCCGTCAAGGTCGCAGACAGCCCGC-ACATCCGCGAG 939

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        940 AAGATCGATGCCAGTGTCGAGTTCCTCCGTACTCAGCTCGACAACAGTGTCTACGGTGTC 999

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db       1000 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCGATCTCGCTCCAAAAG 1059

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db       1060 GCCCTGCTCGAGCACCAGCTCTGCGGTGTCCTCCCCACCTCGATGGATGGCTTTGCGCTC 1119

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db       1120 GGTCGCGGCCTCGAGAACTCGCTTCCGCTCGAAGTCGTCCGAGGCGCGATGACCATCCGT 1179

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db       1180 GTCAACTCGCTCACTCGCGGTCACTCGGCGGTCCGCATCGTCGTCCTCGAAGCCCTCACC 1239

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db       1240 AACTTCCTCAACCACGGCATCACCCCGATCGTCCCGCTTCGAGGCACCATCTCGGCGTCG 1299

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db       1300 GGCGACCTTTCCCCCCTCTCTTACATCGCCGCCTCGATCACCGGCCACCCGGACTCGAAG 1359

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db       1360 GTCCACGT------------CGACGGCAAGATCATGTCCGCCCAGGAGGCGATCGCGCTC 1407

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db       1408 AAGGGTCTTCAGCCCGTCGTCCTCGGTCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 1467

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db       1468 GCCGTCTCCGCCTCGATGGCGACGCTGGCCCTCACCGACGCACACGTCCTCTCGCTCCTC 1527

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db       1528 GCACAGGCGCTCACTGCTCTTACTGTCGAGGCCATGGTCGGACACGCCGGCTCGTTCCAC 1587

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db       1588 CCATTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTGGCGCGCAACATC 1647

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db       1648 CGGACTCTTCTCGAGGGCAGCAAGTACGCCGTCCACCACGAGACTGAAGTCAAGGTCAAG 1707

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1708 GACGACGAGGGCATCCTCAGGCAGGACCGGTACCCGCTCCGCTGCTCGCCGCAGTGGCTC 1767

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1768 GGTCCCCTTGTCAGCGACATGATTCACGCTCACGCTGTCCTCTCGCTCGAGGCTGGTCAG 1827

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1828 TCGACCACCGACAACCCGCTGATCGACCTCGAGAACAAGATGACCCACCATGGCGGAGCC 1887

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1888 TTCATGGCGAGCAGCGTCGGAAACACGATGGAGAAGACTCGCCTCGCCGTCGCGCTGATG 1947

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1948 GGCAAGGTCAGCTTTACTCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGGGCCCTT 2007

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       2008 CCGTCCTGCCTCGCTGCCGAGGACCCTTCCCTCTCTTATCACTGCAAGGGTCTCGACATT 2067

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       2068 GCTGCGGCCGCCTACACTTCCGAGCTCGGTCACCTTGCCAACCCGGTTTCGACCCACGTC 2127

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       2128 CAGCCGGCCGAGATGGGCAACCAGGCCATCAACTCGCTCGCCCTCATCTCGGCCCGCCGC 2187

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       2188 ACCGCCGAGGCGAACGACGTTCTCTCCCTCCTCCTCGCCACCCACCTCTACTGCGTCCTC 2247

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       2248 CAGGCCGTCGACCTCCGCGCGATGGAGTTTGAGCACACCAAGGCGTTCGAGCCGATGGTC 2307

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       2308 ACTGAGCTGTTGAAGCAGCACTTTGGCGC--------------GCTCGCGACGGCCGAA 2352

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       2353 GTCGAGGACAAGGTCCGCAAGTCGATCTACAAGCGGTTGCAGCAGAACAACTCGTACGAC 2412

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       2413 CTCGAGCAGCGGTGGCACGACACGTTCTCGGTCGCGACCGGTGCCGTCGTCGAGGCGCT- 2471

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       2472 ----------CGCCGGCCAGGAGGTCTCGCTCGCGAGCCTCAACGCCTGGAAGGTCGCCT 2521

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       2522 GCGCCGAGAAGGCTATCGCGCTCACGCGCTCCGTCCGCGACTCGTTCTGGGCGGCTCCGT 2581

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       2582 CGTCGTCGTCGCCCGCGCTCAAGTACCTCTCCCCGCGGACGCGCGTCCTGTATTCGTTCG 2641

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2642 TCCGGGAGGAGGTCGGCGTCAAGGCCCGCCGCGGCGATGTCTACCTCGGCAAGCAGGAGG 2701

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2702 TCACGATCGGCACCAACGTCAGCCGCATCTACGAGGCGATCAAGAGCGGTTGCATCGCCC 2761

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2762 CCGTCCTCGTCAAGATGATGGCATAG 2787
```

RESULT 10
US-09-624-693A-14
; Sequence 14, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24

```
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 2311
;   TYPE: DNA
;   ORGANISM: Amanita muscaria
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (18)..(2237)
US-09-624-693A-14

  Query Match             13.0%;  Score 322.6;  DB 4;  Length 2311;
  Best Local Similarity   52.8%;  Pred. No. 3e-51;
  Matches  734;  Conservative    0;  Mismatches  625;  Indels   30;  Gaps    4;

Qy        345 CAGNGTCTACGGNGTCACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGC 404
Db        368 CAGCGTATACGGACTGTCAACTGGTTTCGGTGGCAGTGCTGATACACGGACGGACAAACC 427

Qy        405 NATCTCGCTCCAGAAGGCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNT 464
Db        428 GATGTTGTTGGGGTTTGCCCTTTTGCAACACCAACATGTAGGGATACTGCCCACCTCGAC 487

Qy        465 CGANTCCTTCNGCCTCGGNCGCGGCCTCGAG---------AACTCGCTTCCGCTCGAGGT 515
Db        488 TGAGCCTTTGGACGTCCTACCTCTCCAAGATGCAAATAACACAAGCATGCCAGAGGCGTG 547

Qy        516 CGTCCGCGGCGCCATGACCATCCGCGTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCG 575
Db        548 GATTCGCGGGGCCATTTTGATCCGTATGAATTCGCTAATTCGTGGCCACTCTGGAATCAG 607

Qy        576 CCTCGTCGTCCTCGAGGCGCTCACCAACTTCCTCAACCACGGCATCACCCCCATCGTCCC 635
Db        608 ATGGGAGTTGATCGAAAAGATGAGAGAACTACTCGCGGCCAATGTGATACCTGTCGTTCC 667

Qy        636 CCTCCGCGGCACCATCTCGGCGTCGGGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGC 695
Db        668 CCTGAGAGGCAGCATCTCCTCATCCGGAGATCTGTCTCCCCTATCCTATATCGCAGGCAC 727

Qy        696 CATCACCGGTCACCCGGACNNCAAGGTNCACGTNNTNC------ACGAGGGCANNGAGAA 749
Db        728 GATTATTGGCAACCCATCAATCAAGGTATATCACGGTCCATCAAAGTCCGGAATTCGCCA 787

Qy        750 GATCATGTNCGCCCGCGAGGCGATCGCGCTCTTNGGTCTCGAGCCCGTCGTCCTCGGCCC 809
Db        788 AATTGGATCCTCGAAGGATGTCTTGGCTCTGCATAATATCGAACCTTTCCCACTGGAATC 847

Qy        810 GAAGGAGGGTCTCGGTCTCGTCAACGGCACGGCCGTCTCCGCCTCGATGGCGACCCTCGC 869
Db        848 GAAAGAACCTCTTGGTATTTTGAATGGGACCGCATTCTCGGCATCTGTGGCAGCTTTAGC 907

Qy        870 TCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGGCGCTCACGGCTCTNACGGTCGA 929
Db        908 CCTAAACGAAGCTATCCATCTTGTCTTGTTGGCTCAAGTGTGCACGGCTATGGGGACCGA 967

Qy        930 GGCCATGGTCGGCCACGCCGGCTCGTTCCACCCNTTCCTCCACGACGTCACGCGCCCTCA 989
Db        968 GGCATTGATAGGCACTCGCGCTTCTCATGCACCGTTCATTCATGCCACCGCACGACCACA 1027

Qy        990 CCCGACCCAGATCGAGGTCGCGCGCAACATCCGCACGCTCCTCGAGGGCAGCNNGTTTGC 1049
Db       1028 TCCCGGTCAAGTAGAATGTGCTGAGAACATTTGGAATTTGCTCGATGGGAGTAAATTGGC 1087

Qy       1050 CGTCCACCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGGGCATTCTCCGCCAGGACCG 1109
Db       1088 TCAGTTAGAAGAGCACGAAGTTCGCCTAGAAGACGATAAATACACCCTTCGGCAGGACCG 1147

Qy       1110 CTACCCGCTCCGCACGTCGCCTCAGTGGCTCGGCCCGCTCGTCAGCGACATGATTCACGC 1169
Db       1148 TTATCCACTCCGAACTTCGCCTCAATTCCTTGGGCCTCAGATTGAAGACATAATCTCCGC 1207

Qy       1170 CCACGCNGTCCTCTCGCTCGAGGCCGGNCAGT---CGACGACCGACAACCCGCTCATCGA 1226
Db       1208 TTTCCAGACTGTAACGCAGGAGTGTAATTACTTACCAGCTACTGACAATCCACTGATTGA 1267

Qy       1227 CGTCGAGAACAAGANGACCCACCACGGCGGCAACTTCCAGGCGNCCGCTGTCGCNAACAC 1286
Db       1268 TGGGGAGACTGGCGAATCTCACCACGGTGGCAATTTCCAAGCGATGGCTGTAACTAATGC 1327

Qy       1287 GATGGAGAAGACTCGCCTCGCNCTCGCCCTGATCGGCAAGCTCAACTTCACGCAGCTCAC 1346
Db       1328 AATGGAGAAGACGCGACTTGCTTTACATCACGTTGGCAAATTACTATTTCCCAGAGCAC 1387

Qy       1347 CGAGATGCTCAACGCCGGCATGAACCGCGGCCTNCCNTCCTGCCTCGCTGCCGAGGACCC 1406
Db       1388 TGAATTAGTCAATCCTGCGATGAACCGCGGTCTGCCGCCTTCAGTAGCTGCCACAGATCC 1447

Qy       1407 NTCGCTCTCCTATCACTGCAAGGGCCTCGACATTGCCGCNGCNGCNTACACTTCGGAGCT 1466
Db       1448 ATCTCTCAACTACCACGCCAAAGGACTAGACATAGCAACTGCGGCCTACGTAGCCGA--- 1504

Qy       1467 CGGNCACCTTGCCAACCCGGTNACGACCCACGTCCAGCCGGCNGAGATGGGCAACCAGGC 1526
Db       1505 ---------AGCGACTCCTGGCCCCACTCACATTCAGTCGGCAGAAATGCACAACCAAGC 1555

Qy       1527 CGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGCACNGCCGAGGCCAACGACGTCCTTTC 1586
Db       1556 TGTTAACTCCCTGGCGTTGATTTCTGCTCGGGCTACCATCACATCGTTGGAAGTGCTAAC 1615

Qy       1587 TCTCCTCCTCGCCACCCACCTCTACTGCGTNCTCCAGGCCGTCGACCTCCGCGCGATGGA 1646
Db       1616 ATCTCTGATCGCGTCTTACTTGTATATTCTATGCCAAGCTCTCGACCTCCGTGCCCTTCA 1675

Qy       1647 GTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTCNCNNCGCTCNTCNAGCAGCACTTTGG 1706
Db       1676 GCGCGAGTTCTTGCCCGGTCTAGACATCATCATTCGTGAGGAGTTAAGATCGTCATTTGG 1735

Qy       1707 CNCNGCCCT 1715
Db       1736 ATCTTTCCT 1744

RESULT 11
US-09-615-192A-245
; Sequence 245, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
;  APPLICANT:  Bloksberg, Leonard N.
;  APPLICANT:  Havukkala, Ilkka
;  TITLE OF INVENTION:  Materials and Methods for the
;  TITLE OF INVENTION:  Modification of Plant Lignin Content
;  FILE REFERENCE:  11000.1003c4U
;  CURRENT APPLICATION NUMBER: US/09/615,192A
;  CURRENT FILING DATE:  2000-07-12
;  PRIOR APPLICATION NUMBER:  US 08/975,316
;  PRIOR FILING DATE:  1997-11-21
;  PRIOR APPLICATION NUMBER:  US 08/713,000
;  PRIOR FILING DATE:  1996-09-11
;  PRIOR APPLICATION NUMBER:  US 09/169,789
;  PRIOR FILING DATE:  1998-10-09
;  NUMBER OF SEQ ID NOS:  405
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 245
;   LENGTH: 1455
;   TYPE: DNA
;   ORGANISM: Eucalyptus grandis
US-09-615-192A-245

  Query Match              8.6%;  Score 213.2;  DB 4;  Length 1455;
  Best Local Similarity   54.8%;  Pred. No. 4.4e-31;
  Matches  480;  Conservative    0;  Mismatches  384;  Indels   12;  Gaps    4;

Qy        465 CGANTCCTTCNGCCTCGGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGG 524
Db        592 CGGGATCTTCGGCAACGGCACGGAGTCGTGCCACACCCTGCCTCAATCCTCCACCCGAGC 651
```

```
Qy    525 CGCCATGACCATCCGCGTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGT 584
Db    652 CGCCATGCTCGTCCGGGTCAACACCCTCCTCCAGGGCTACTCCGGCATCCGTTTTGAGAT 711

Qy    585 CCTCGAGGCGCTCACCAACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGG 644
Db    712 CCTCGAGGCCATCACCAAGTTCCTCAACCACAACATCACCCCGTGCCTGCCCCTCAGGGG 771

Qy    645 CACCATCTCGGCGTCGGGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGG 704
Db    772 CACCATCACTGCCTCAGGCGACTTGGTCCCCCTCTCCTACATTGCCGGGCTCCTGACGGG 831

Qy    705 TCACCCGGACNNCAAGGTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCG 764
Db    832 CCGGCCCAACTCCAAGG---CCGTCGGGCCTGATGGGAAGTCCCTGGACGCTGTCGAGGC 888

Qy    765 CGAGGCGATCGCGCTCTTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGG 824
Db    889 CTTCCGGCTCGCCGGGATTGACACGGGCTTCTTCGAGCTGCAGCCAAAGGAAGGGTTGGC 948

Qy    825 TCTCGTCAACGGCACGGCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACA 884
Db    949 GCTCGTGAACGGCACGGCAGTCGGGTCTGGCCTGGCTTCCATCGTCCTCTTCGAGGCCAA 1008

Qy    885 CATGCTCTCGCTCCTCTCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCA 944
Db   1009 CATACTCGCGGTCCTGTCCGAGGTCCTGTCAGCGATCTTCGCAGAGGTGATGCAGGGGAA 1068

Qy    945 CGCCGGCTCGTTCCACCCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGA 1004
Db   1069 GCCGGAGTTCACAGACCACTTGACGCATAAATTGAAGCACCATCCCGGG---CAGATTGA 1125

Qy   1005 GGTCGCGCGCAACATCCGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGA 1064
Db   1126 GTCTGCGGCTATAATGGAGCACATTTTGGATGGAAGCGC---TTACGTGAAGGCTGCTAA 1182

Qy   1065 GGAGGTCAAGGTCAAGGACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCAC 1124
Db   1183 AAAGTTGCACGAGATGGATCCGCTCCAGAAGCCAAAGCAGGACAGGTACGCTCTCAGGAC 1242

Qy   1125 GTCGCCTCAGTGGCTCGGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTC 1184
Db   1243 TTCTCCCCAGTGGCTAGGGCC---CCAGATTGAGGTGATCCGAGCGGCAACCAAGATGAT 1299

Qy   1185 GCTCGAGGCCGGNCAGTCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGAC 1244
Db   1300 TGAGAGGGAAATCAATTCGGTCAATGACAACCCGCTGATCGATGTCGCGAGGAACAAGGC 1359

Qy   1245 CCACCACGGCGGCAACTTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCT 1304
Db   1360 CCTGCACGGTGGGAACTTCCAGGGGACCCCGATTGGTGTCTCCATGGACAACACTCGCCT 1419

Qy   1305 CGCNCTCGCCCTGATCGGCAAGCTCAACTTCACGCA 1340
Db   1420 GGCGGTTGCGTCCATAGGGAAGCTCATGTTCGCGCA 1455
```

RESULT 12
US-09-615-192A-247
; Sequence 247, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 247
; LENGTH: 1520
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-615-192A-247

Query Match 6.9%; Score 170.2; DB 4; Length 1520;
Best Local Similarity 49.7%; Pred. No. 3.8e-23;
Matches 480; Conservative 0; Mismatches 473; Indels 13; Gaps 4;

```
Qy    791 AGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACGGCCGTCTCCG 850
Db    124 AGCCCTTTGAATTGCAGCCTAAAGAAGGTCTGGCCATTGTCAATGGCACTTCAGTGGGAG 183

Qy    851 CCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCAGGCGC 910
Db    184 CAGCACTGGCTTCCATTGTGTGTTTCGATGCCAATGTTCTTGCTCTGCTCTCTGAAGTAA 243

Qy    911 TCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCACCCNTTCCTCC 970
Db    244 TCTCTGCCATGTTCTGCGAGGTTATGAATGGGAAGCCTGAGTTTACAGATCCATTAACTC 303

Qy    971 ACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATCCGCACGCTCC 1030
Db    304 ACAAGCTGAAGCACCAT---CCTGGCCAAATGGAAGCTGCAGCGATCATGGAGTATGTCT 360

Qy   1031 TCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAGGACGACGAGG 1090
Db    361 TGGACGGGA----GTCTTATATGAAACACGCTGCTAAGCTCCATGAGATGAATCCTCTGC 416

Qy   1091 GCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTCGGCCCGCTCG 1150
Db    417 AGAAGCCAAAGCAGGATCGCTATGCGCTTCGCACTTCGCCTCAGTGGCTCGGCCCTCAGG 476

Qy   1151 TCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAGTCGACGACCG 1210
Db    477 T---GGAGATTATCAGATCTGCAACTCACATGATTGAGCGGGAAATCAATTCTGTGAATG 533

Qy   1211 ACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAACTTCCAGGCGN 1270
Db    534 ACAATCCAGTAATTGATGTTGCCAGAGACAAAGCTCTACATGGAGGGAATTTCCAGGGCA 593

Qy   1271 CCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATCGGCAAGCTCA 1330
Db    594 CACCTATTGGTGTTTCCATGGATAATCTTCGTCTGTCAATTTCAGCAATTGGGAAATTGA 653

Qy   1331 ACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTNCCNTCCTGCC 1390
Db    654 TGTTCGCTCAATTCTCAGAGCTTGTGAATGATTACTACAATGGAGGCTTGCCTTCGAATC 713

Qy   1391 TCGCTGCCGAG---GACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATTGCCGCNG 1447
Db    714 TGAGTGGTGGGCCTAATCCCAGCCTGGATTATGGACTGAAAGGGGCCGAGATCGCTATGG 773

Qy   1448 CNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTCCAGCCGG 1507
Db    774 CTTCTTACACTTCTGAGCTTCTTTACCTGGCAAATCCTGTCACCAGCCATGTACAGAGCG 833

Qy   1508 CNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGCACNGCCG 1567
Db    834 CCGAACAGCATAACCAGGATGTCAATTCTCTGGGTCTCGTTTCAGCTAGAAAATCTGCCG 893

Qy   1568 AGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTCCAGGCCG 1627
Db    894 AGGCCATCGATATTCTGAAGCTGATGCTCTCCACATACCTGACAGCTCTGTGCCAGGCTG 953

Qy   1628 TCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTCNCNNCGC 1687
Db    954 TGGATTTAAGGCATCTGGAGGAAAACATGCTGGCCACTGTGAAGCAGATTGTTTCTCAGG 1013
```

```
Qy       1688 TCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAACTCGNGG 1747
Db       1014 TAGCCAAGAAAACCCTGAGCACAGGGCTCAACGGGGAGCTTTTGCCAGGCCGTTTCTGCG 1073

Qy       1748 ACAAGG 1753
Db       1074 AAAAGG 1079
```

RESULT 13
US-09-252-991A-16475
; Sequence 16475, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16475
;   LENGTH: 1548
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16475

```
  Query Match             5.3%;  Score 131.6;  DB 4;  Length 1548;
  Best Local Similarity   46.5%;  Pred. No. 5e-16;
  Matches  572;  Conservative    0;  Mismatches  606;  Indels   52;  Gaps    5;

Qy        471 CTTCNGCCTCGGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCAT 530
Db        336 CCTGTCCCATGCCGCCGGCGTCGGCGAGGCGCTGGACGACGCCATGGTGCGGCTGGTCAT 395

Qy        531 GACCATCCGCGTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGA 590
Db        396 GCTGCTCAAGGTGAACAGCCTGGCGCGCGGGTTCTCCGGCATCCGCCGCAAGGTGATCGA 455

Qy        591 GGCGCTCACCAACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCAT 650
Db        456 CGCGCTGATCGCGCTGATCAACGCCGAGGTCTATCCGCACATCCCGCTGAAAGGCTCGGT 515

Qy        651 CTCGGCGTCGGGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCC 710
Db        516 GGGTGCGTCCGGCGACCTGGCGCCGCTGGCGCACATGTCGCTGGTGCTGATCGGCGAAAG 575

Qy        711 GGACNNCAAGGTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGC 770
Db        576 CCGGGCGCGCCATCGCGGTG------------------AGTGGCTGCCGGCCGCCGAAGC 617

Qy        771 GATCGCGCTCTTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGT 830
Db        618 GCTGGCGGTGGCCGGGCTGGAGCCGCTGACCCTGGCGGCGAAGGAAGGCCTGGCCCTGCT 677

Qy        831 CAACGGCACGGCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCT 890
Db        678 CAATGGCACCCAGGTGTCCACCGCCTACGCGCTGCGCGGGTTGTTCGAGGCCGAGGACCT 737

Qy        891 CTCGCTCCTCTCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGG 950
Db        738 GTTCGCCGCCGCCACCGTCTGCGGCGGGCTCAGCGTCGAGGCCATGCTCGGTTCGCGGGC 797

Qy        951 CTCGTTCCACCCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGC 1010
Db        798 GCCGTTCGATGCGCGCATCCATGCCGCGCGCGGCCAGCGTGGGCAGATCGACGTGGCCGC 857

Qy       1011 GCGCAACATCCGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGT 1070
Db        858 GGCCTATCGCGACCTGCTCACCGCCAGCAGCGAGGTGGCGCGCTCCCATGAGAAGTGCGA 917

Qy       1071 CAAGGTCAAGGACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCC 1130
Db        918 CAAGGT-----------------------CCAGGACCCCTATTCGCTGCGTTGCCAGCC 953

Qy       1131 TCAGTGGCTCGGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGA 1190
Db        954 GCAGGTGATGGGCGCCTGCCTGACCCAGATGCGCCAGGCCGCCGAGGTGCTGGAGATCGA 1013

Qy       1191 GGCCGGNCAGTCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCA 1250
Db       1014 AGCC---AACGCGGTGTCCGACAACCCGCTGGTATTCGCCGCCGAGGGCGACGTGATCTC 1070

Qy       1251 CGGCGGCAACTTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCT 1310
Db       1071 CGGCGGCAACTTCCACGCCGAACCGGTGGCGATGGCCGCCGACAACCTGGCCCTGGCGTT 1130

Qy       1311 CGCCCTGATCGGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAA 1370
Db       1131 GGCCGAGATCGGTTCGCTGTCGGAACGGCGCATCTCGCTGATGATGGACATGCACATGTC 1190

Qy       1371 CCGCGGCCTNCCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGG 1430
Db       1191 GCAGTTGCCGCCGTTCCTGGTGGCCAACGGCG------GGGTCAACTCCGGCTTCATGAT 1244

Qy       1431 CCTCGACATTGCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNAC 1490
Db       1245 CGCCCAGGTCACCGCCGCCGCGCTGGCCAGCGACAACAAGGCCCTGGCCCATCCGGCCAG 1304

Qy       1491 GACCCACGTCCAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTC 1550
Db       1305 CGTCGACAGCCTGCCGACCTCGGCCAACCAGGAAGACCATGTGTCGATGGCGCCGAACGC 1364

Qy       1551 GGCNCGCCGCACNGCCGAGGCCAACGACGTCCTTTCTC-TCCTCCTCGCCACCCACCTCT 1609
Db       1365 CGGCAAACGGCTCTGGGCGATGGCCGAGAACGTCCCGCGGGATTCTCGCCGTGGAGTGGC 1424

Qy       1610 ACTGCGTNCTCCAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCG 1669
Db       1425 TGGGCGCCTGCCAGGGCCTGGACTTCCGCGAGGGCCTGAAGAGTTCGCCGAAGCTGGAGC 1484

Qy       1670 ACCCGNTNNTCNCNNCGCTCNTCNAGCAGC 1699
Db       1485 AGGCTCGCCGCCTGCTGCGCGACAAGGTGC 1514
```

RESULT 14
US-09-252-991A-16126/c
; Sequence 16126, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16126
;   LENGTH: 2295
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16126

```
  Query Match             5.3%;  Score 131.6;  DB 4;  Length 2295;
  Best Local Similarity   46.5%;  Pred. No. 5.2e-16;
  Matches  572;  Conservative    0;  Mismatches  606;  Indels   52;  Gaps    5;
```

```
Qy        471 CTTCNGCCTCGGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCAT 530
Db       1183 CCTGTCCCATGCCGCCGGCGTCGGCGAGGCGCTGGACGACGCCATGGTGCGGCTGGTCAT 1124

Qy        531 GACCATCCGCGTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGA 590
Db       1123 GCTGCTCAAGGTGAACAGCCTGGCGCGCGGGTTCTCCGGCATCCGCCGCAAGGTGATCGA 1064

Qy        591 GGCGCTCACCAACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCAT 650
Db       1063 CGCGCTGATCGCGCTGATCAACGCCGAGGTCTATCCGCACATCCCGCTGAAAGGCTCGGT 1004

Qy        651 CTCGGCGTCGGGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCC 710
Db       1003 GGGTGCGTCCGGCGACCTGGCGCCGCTGGCGCACATGTCGCTGGTGCTGATCGGCGAAAG 944

Qy        711 GGACNNCAAGGTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGC 770
Db        943 CCGGGCGCGCCATCGCGGTG------------------AGTGGCTGCCGGCCGCCGAAGC 902

Qy        771 GATCGCGCTCTTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGT 830
Db        901 GCTGGCGGTGGCCGGGCTGGAGCCGCTGACCCTGGCGGCGAAGGAAGGCCTGGCCCTGCT 842

Qy        831 CAACGGCACGGCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCT 890
Db        841 CAATGGCACCCAGGTGTCCACCGCCTACGCGCTGCGCGGGTTGTTCGAGGCCGAGGACCT 782

Qy        891 CTCGCTCCTCTCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGG 950
Db        781 GTTCGCCGCCGCCACCGTCTGCGGCGGGCTCAGCGTCGAGGCCATGCTCGGTTCGCGGGC 722

Qy        951 CTCGTTCCACCCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGC 1010
Db        721 GCCGTTCGATGCGCGCATCCATGCCGCGCGCGGCCAGCGTGGGCAGATCGACGTGGCCGC 662

Qy       1011 GCGCAACATCCGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGT 1070
Db        661 GGCCTATCGCGACCTGCTCACCGCCAGCAGCGAGGTGGCGCGCTCCCATGAGAAGTGCGA 602

Qy       1071 CAAGGTCAAGGACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCC 1130
Db        601 CAAGGT------------------------CCAGGACCCCTATTCGCTGCGTTGCCAGCC 566

Qy       1131 TCAGTGGCTCGGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGA 1190
Db        565 GCAGGTGATGGGCGCCTGCCTGACCCAGATGCGCCAGGCCGCCGAGGTGCTGGAGATCGA 506

Qy       1191 GGCCGGNCAGTCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCA 1250
Db        505 AGCC---AACGCGGTGTCCGACAACCCGCTGGTATTCGCCGCCGAGGGCGACGTGATCTC 449

Qy       1251 CGGCGGCAACTTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCT 1310
Db        448 CGGCGGCAACTTCCACGCCGAACCGGTGGCGATGGCCGCCGACAACCTGGCCCTGGCGTT 389

Qy       1311 CGCCCTGATCGGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAA 1370
Db        388 GGCCGAGATCGGTTCGCTGTCGGAACGGCGCATCTCGCTGATGATGGACATGCACATGTC 329

Qy       1371 CCGCGGCCTNCCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGG 1430
Db        328 GCAGTTGCCGCCGTTCCTGGTGGCCAACGGCG------GGGTCAACTCCGGCTTCATGAT 275

Qy       1431 CCTCGACATTGCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNAC 1490
Db        274 CGCCCAGGTCACCGCCGCCGCGCTGGCCAGCGACAACAAGGCCCTGGCCCATCCGGCCAG 215

Qy       1491 GACCCACGTCCAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTC 1550
Db        214 CGTCGACAGCCTGCCGACCTCGGCCAACCAGGAAGACCATGTGTCGATGGCGCCGAACGC 155

Qy       1551 GGCNCGCCGCACNGCCGAGGCCAACGACGTCCTTTCTC-TCCTCCTCGCCACCCACCTCT 1609
Db        154 CGGCAAACGGCTCTGGGCGATGGCCGAGAACGTCCCGCGGGATTCTCGCCGTGGAGTGGC 95

Qy       1610 ACTGCGTNCTCCAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCG 1669
Db         94 TGGGCGCCTGCCAGGGCCTGGACTTCCGCGAGGGCCTGAAGAGTTCGCCGAAGCTGGAGC 35

Qy       1670 ACCCGNTNNTCNCNNCGCTCNTCNAGCAGC 1699
Db         34 AGGCTCGCCGCCTGCTGCGCGACAAGGTGC 5
```

RESULT 15
US-09-615-192A-97
; Sequence 97, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; TITLE OF INVENTION: Modification of Plant Lignin Content
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 97
;   LENGTH: 577
;   TYPE: DNA
;   ORGANISM: Pinus radiata
US-09-615-192A-97

```
  Query Match             4.5%;  Score 111;  DB 4;  Length 577;
  Best Local Similarity   54.0%;  Pred. No. 2.8e-12;
  Matches  285;  Conservative    0;  Mismatches  234;  Indels    9;  Gaps    3;

Qy        786 TCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACGGCCGT 845
Db         52 TCTGGAACCGTTTAAACTGCAGCCCAAGGAAGGACTGGCTCTCGTCAACGGCACAGCGGT 111

Qy        846 CTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTCTCGCA 905
Db        112 GGGATCCGCCGTGGCCGCGTCCGTCTGTTTTGACGCCAACGTGCTGGGCGTGCTGGCTGA 171

Qy        906 GGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCACCCNTT 965
Db        172 GATTCTGTCTGCGCTCTTCTGCGAGGTGATGCAAGGGAAACCGGAGTTCGTAGATCCGTT 231

Qy        966 CCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATCCGCAC 1025
Db        232 AACCCACCAGTTGAAGCAC---CACCCAGGGCAGATCGAAGCCGCGGCCGTCATGGAGTT 288

Qy       1026 GCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAGGACGA 1085
Db        289 CCTCCTCGACGGTAGCGACT---ACGTGAAAGAAGCAGCGCGGCTTCACGAGAAAGACCC 345

Qy       1086 CGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTCGGCCC 1145
Db        346 GTTGAGCAAACCGAAACAAGACCGCTACGCTCTGCGAACATCGCCACAGTGGTTGGGGCC 405

Qy       1146 GCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAGTCGAC 1205
Db        406 TCCGATCGAAGTCATCCGCGCTGCTACTCACTCCATCGAGCGGGAGATC---AATTCCGT 462

Qy       1206 GACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAACTTCCA 1265
```

```
Db        463 CAACGACAATCCGTTAATCGATGTCTCCAGGGACATGGCTCTCCACGGCGGCAACTTCCA 522

Qy       1266 GGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGC 1313
              ||   |  |  |||      | |||||  || |  ||  || |  | ||
Db        523 GGGAACACCCATCGGAGTTTCCATGGACAACATGCGAATCTCTTTGGC 570
```

Search completed: September 12, 2004, 04:26:57
Job time : 199.6 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        September 11, 2004, 23:12:54 ; Search time 1606.68 Seconds
                                          (without alignments)
                                          7750.794 Million cell updates/sec

Title:         US-09-939-408A-20
Perfect score: 2475
Sequence:      1 atggccccctccntcgactc..........nnnnnnnnnnnnnnnnnnn 2475

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters:      6608766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
               1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
               2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
               3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
               4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
               5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
               6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
               7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
               8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
               9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
               10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
               11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
               12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
               13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
               14:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
               15:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
               16:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
               17:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
               18:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
               19:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2128	86.0	2475	9	US-09-939-408A-20	Sequence 20, Appl
2	1944	78.5	2163	9	US-09-939-408A-29	Sequence 29, Appl
3	1699.4	68.7	2439	9	US-09-939-408A-18	Sequence 18, Appl
4	1629	65.8	2419	9	US-09-939-408A-12	Sequence 12, Appl
5	1626.4	65.7	2151	9	US-09-765-873A-9	Sequence 9, Appli
6	1626.4	65.7	2151	15	US-10-138-970A-23	Sequence 23, Appl
7	1626.4	65.7	2151	15	US-10-188-523B-9	Sequence 9, Appli
8	1626.4	65.7	2151	16	US-10-374-366-3	Sequence 3, Appli
9	1624.8	65.6	2151	16	US-10-374-366-9	Sequence 9, Appli
10	1624.8	65.6	2151	16	US-10-374-366-13	Sequence 13, Appl
11	1624.8	65.6	2151	16	US-10-374-366-23	Sequence 23, Appl
12	1623.2	65.6	2151	9	US-09-765-873A-7	Sequence 7, Appli
13	1623.2	65.6	2151	15	US-10-188-523B-7	Sequence 7, Appli
14	1623.2	65.6	2151	16	US-10-374-366-1	Sequence 1, Appli
15	1623.2	65.6	2151	16	US-10-374-366-5	Sequence 5, Appli
16	1623.2	65.6	2151	16	US-10-374-366-7	Sequence 7, Appli
17	1623.2	65.6	2151	16	US-10-374-366-15	Sequence 15, Appl
18	1623.2	65.6	2151	16	US-10-374-366-21	Sequence 21, Appl
19	1621.6	65.5	2151	16	US-10-374-366-11	Sequence 11, Appl
20	1621.6	65.5	2151	16	US-10-374-366-17	Sequence 17, Appl
21	1618.4	65.4	2151	16	US-10-374-366-19	Sequence 19, Appl
22	1610.4	65.1	2061	9	US-09-765-873A-31	Sequence 31, Appl
23	1610.4	65.1	2061	15	US-10-188-523B-31	Sequence 31, Appl
24	1598.6	64.6	2787	9	US-09-939-408A-16	Sequence 16, Appl
25	1213.6	49.0	2703	16	US-10-439-478-1	Sequence 1, Appli
26	1124.4	45.4	2741	9	US-09-939-408A-28	Sequence 28, Appl
27	757.4	30.6	2465	17	US-10-439-479-24	Sequence 24, Appl
28	730.8	29.5	1020	15	US-10-138-970A-3	Sequence 3, Appli
29	471.4	19.0	2539	13	US-10-425-114-26376	Sequence 26376, A
30	463.2	18.7	2454	13	US-10-425-114-34124	Sequence 34124, A
31	462.2	18.7	2468	13	US-10-425-114-30888	Sequence 30888, A
32	462.2	18.7	2469	13	US-10-425-114-31903	Sequence 31903, A
33	461.2	18.6	2417	13	US-10-425-114-31197	Sequence 31197, A
34	455.8	18.4	2554	17	US-10-767-701-15815	Sequence 15815, A
35	450.2	18.2	2972	17	US-10-437-963-48311	Sequence 48311, A
36	442.2	17.9	2544	17	US-10-437-963-55261	Sequence 55261, A
37	430.2	17.4	2325	17	US-10-767-701-15813	Sequence 15813, A
38	425.6	17.2	2079	13	US-10-425-114-1334	Sequence 1334, Ap
39	411.2	16.6	1092	17	US-10-439-479-9	Sequence 9, Appli
40	400.4	16.2	2329	17	US-10-437-963-41428	Sequence 41428, A
41	392.8	15.9	2354	13	US-10-425-114-28070	Sequence 28070, A
42	380.8	15.4	2774	17	US-10-437-963-38582	Sequence 38582, A
43	379.2	15.3	2340	13	US-10-425-114-2857	Sequence 2857, Ap
44	365.8	14.8	2145	17	US-10-437-963-40094	Sequence 40094, A
45	360.6	14.6	2007	13	US-10-425-114-2110	Sequence 2110, Ap

ALIGNMENTS

```
RESULT 1
US-09-939-408A-20
; Sequence 20, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 2475
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: misc_difference
;   LOCATION: (13)..(2475)
;   OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus
;   OTHER INFORMATION: Sequence of SEQ ID NOs: 12, 16, and 18
US-09-939-408A-20

  Query Match             86.0%;  Score 2128;  DB 9;  Length 2475;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2408;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
```

```
Qy      2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db      2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280

Qy      2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db      2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340

Qy      2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db      2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400

Qy      2401 NNNANAAA 2408
Db      2401 NNNANAAA 2408


RESULT 2
US-09-939-408A-29
; Sequence 29, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 2163
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:   Preferred
;   OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
;   NAME/KEY: CDS
;   LOCATION: (1)..(2163)
;   OTHER INFORMATION:
US-09-939-408A-29

  Query Match            78.5%;  Score 1944;  DB 9;  Length 2163;
  Best Local Similarity   93.2%;  Pred. No. 0;
  Matches 2037;  Conservative    0;  Mismatches  126;  Indels   23;  Gaps    7;

Qy         1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db         1 ATGGCCCCCTCCBTCGACTCGATCGCGACCTCGBTCGCCAACGGCBTCDCVAACGGVTHG 60

Qy        61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db        61 CACGCCGCTCCGHCCAAGYCGSCWACGGGCGCCACGTCCACHCTCMGMCGGCCGBC-CDG 119

Qy       121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db       120 GCTCGCTCCTCCCG-CCACCCAGVHGACGCAGCTCGACATCGTBGAGVAGATCCTCGCCG 178

Qy       181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db       179 ACCCCACCGAC---GACGWSVTCGAACTCGACGGGTACACCCTCACCCTCGGHGACGTCG 235

Qy       241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db       236 TCGGCGCCGCBCGCAAGGGCCGCHCBGTCCGCGTC-CAGACAGMCGACGAGATCCGCGCA 294

Qy       301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db       295 AAGATCGACAAVAGBGTCGAGTTCCTCCGBDCBCAGCTCBACAACAGBGTCTACGGHGTC 354

Qy       361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db       355 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCVATCTCGCTCCAGAAG 414

Qy       421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db       415 GCBCTCCTCGAGCACCAGCTCTGCGGTGTBCTCCCBACGTCGDTCGABTCCTTCVGCCTC 474

Qy       481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db       475 GGHCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 534

Qy       541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db       535 GTCAACTCGCTCACBCGCGGCCACTCGGCBGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 594

Qy       601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db       595 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 654

Qy       661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db       655 GGCGACCTCTCCCCBCTCTCBTACATCGCCGCCGCCATCACCGGTCACCCGGACDBCAAG 714

Qy       721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db       715 GTBCACGTYKTSCACGAGGGCAMSGAGAAGATCATGTHCGCCCGCGAGGCGATCGCGCTC 774

Qy       781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db       775 TTBGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 834

Qy       841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db       835 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 894

Qy       901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db       895 TCGCAGGCGCTCACGGCTCTBACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 954

Qy       961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db       955 CCVTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1014

Qy      1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db      1015 CGCACGCTCCTCGAGGGCAGCHVGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1074

Qy      1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db      1075 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1134

Qy      1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db      1135 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCBGTCCTCTCGCTCGAGGCCG---AG 1191

Qy      1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db      1192 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGAHGACCCACCACGGCGGCAAC 1251

Qy      1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db      1252 TTCCAGGCGDCCGCTGTCGCVAACACGATGGAGAAGACTCGCCTCGCVCTCGCCCTGATC 1311

Qy      1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db      1312 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTB 1371

Qy      1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
```

```
Db     1372 CCBTCCTGCCTCGCTGCCGAGGACCCBTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1431

Qy     1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db     1432 GCCGCBGCBGCBTACACTTCGGAGCTCGGHCACCTTGCCAACCCGGTBACGACCCACGTC 1491

Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db     1492 CAGCCGGCHGAGATGGGCAACCAGGCCGTCAACTCGCTCGCBCTCATCTCGGCBCGCCGC 1551

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db     1552 ACBGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTBCTC 1611

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db     1612 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGVTBVTC 1671

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db     1672 VCBDCGCTCHTCVAGCAGCACTTTGGCDCYGCCCTCGACGGC------------WACGAA 1719

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db     1720 CTCGHGGACAAGGTCAACAAGDCGCTCDACAAGCGBCTCGAGCAGACCAACTCGTACGAC 1779

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db     1780 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGCTCCTC 1839

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db     1840 TCGTCCTCGCCBYCTGCCA--AGGTCTCGCTCGCBGCCGTCAACGCCTGGAAGGTCGCCT 1897

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db     1898 CCGCCGAGAAGGCCATCTCGCTCACGCGCBAVGTCCGCGACHCCTTCTGGBCGGCBCCGT 1957

Qy     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db     1958 CGTCGTCGTCGCCCGCGCTCDCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2017

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db     2018 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2077

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db     2078 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGDVCGGCHGCATCAACC 2137

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db     2138 ACGTCCTCGTCAAGATGCTCGCDTAG 2163
```

RESULT 3
US-09-939-408A-18
; Sequence 18, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Rhodotorula toruloides
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2148)
US-09-939-408A-18

Query Match 68.7%; Score 1699.4; DB 9; Length 2439;
Best Local Similarity 79.4%; Pred. No. 0;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

```
Qy        1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db        1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy       61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db       61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy      121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db      101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy      181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db      161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy      241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db      218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy      301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db      277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy      361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db      337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy      421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db      397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy      481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db      457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy      541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db      517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy      601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db      577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy      661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db      637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy      721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db      697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db      757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db      817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876
```

```
Qy   901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db   877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy   961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db   937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy  1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db   997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy  1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db  1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy  1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db  1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy  1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db  1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy  1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db  1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy  1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db  1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy  1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db  1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy  1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db  1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy  1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db  1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy  1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db  1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy  1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db  1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy  1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740
Db  1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy  1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db  1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy  1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db  1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy  1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db  1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy  1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db  1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy  1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db  1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy  2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db  2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy  2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db  2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy  2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db  2126 ACGTCCTCCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCCATACCC 2185

Qy  2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNNN 2280
Db  2186 TATCCCGCCTGCACTCTTAGGACTCGCTTCTTGTCGGACTCGGATCTCGCATCGCTTCTT 2245

Qy  2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db  2246 TCGTTCTTGGCTGCCTCTC-TAGACCGTGTCCGTATTACCTCGAGATTGTGAATACAAGC 2304

Qy  2341 NGTNNNCANNNACNCNTNNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db  2305 AGTACCCATCCACGCATCCGATAAATCAGGGAGAGAATCTACGCTTGCGGGAGCTTCTTG 2364

Qy  2401 NNNANAAA 2408
Db  2365 CGCATAAA 2372
```

RESULT 4
US-09-939-408A-12
; Sequence 12, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 2419
; TYPE: DNA
; ORGANISM: Rhodotorula graminis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)..(2196)
; NAME/KEY: modified_base
; LOCATION: (494)
; OTHER INFORMATION: Other information: y = t or c
; NAME/KEY: unsure
; LOCATION: (493)..(495)
; OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408A-12

```
Query Match            65.8%;  Score 1629;  DB 9;  Length 2419;
Best Local Similarity  78.0%;  Pred. No. 0;
Matches 1878;  Conservative    1;  Mismatches  504;  Indels   25;  Gaps    8;

Qy     1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
```

```
Db     37 ATGGCCCCTTCCTTGGACTCGCTCGCCACCACGCTCGCCAACGGCTTTACCAACGGCTCG 96

Qy     61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db     97 CACGCCGCTCCGACCAAGTCGGCTGCGGGCCCCACTTCGGCTCTCCGCCGCACGC--CCG 154

Qy    121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db    155 GCCTCGATGGCCACGCCGCGCACCAGTCGCAGCTCGAGATCGTGCAGGAGCTCCTCAGCG 214

Qy    181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db    215 ACCCCACCGAC---GACGTCGTCGAGCTCAGCGGGTACAGCCTCACCGTCCGTGACGTTG 271

Qy    241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db    272 TCGGCGCCGCCCGCAAGGGGCGCAGGGTCCGCGTC-CAGAACGACGACGAGATCCGCGCA 330

Qy    301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db    331 CGCGTCGACAAGAGCGTCGACTTCCTCAAGGCCCAGCTTCAGAACTCGGTCTACGGAGTC 390

Qy    361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db    391 ACCACGGGTTTCGGTGGCTCGGCCGACACGAGGACTGAGGATGCAGTCAGCCTCCAGAAG 450

Qy    421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db    451 GCGCTCATCGAGCACCAGCTCTGCGGCGTGACGCCGACGTCCGYCTCGTCCTTCAGCGTC 510

Qy    481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db    511 GGACGCGGCCTCGAGAACACGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGGTCATCCGC 570

Qy    541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db    571 GTCAACTCGCTCACGCGTGGCCACTCGGCCGTCCGCCTCGTCGTCCTTGAGGCGCTCACC 630

Qy    601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db    631 AACTTCTTGAACCACCGCATCACGCCCATCGTCCCCCTCCGCGGCTCCATCTCGGCGTCG 690

Qy    661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db    691 GGCGACCTCAGCCCGCTCTCGTACATCGCCGGCGCCATCACCGGTCACCCCGACGTCAAG 750

Qy    721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db    751 GTTCACGTTTTGCACGAGGGAACCGAGAAGATCATGTTTGCGCGCGAGGCCATCTCGCTC 810

Qy    781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db    811 TTTGGTCTCGAGGCAGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTGGTCAACGGAACG 870

Qy    841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db    871 GCCGTCTCCGCCTCGATGGCGACCCTCAGTCTGCACGACTCGCACATGCTCTCGCTCCTC 930

Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    931 TCGCAGGCCTTGACGGCTCTCACGGTGGAGGCCATGGTCGGCCAGCAGGGCTCGTTCGCG 990

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    991 CCGTTCATCCACGACGTCTGCGCCCCGCACCCCGGCCAGGTCGAGGTCGCGCGCAACATC 1050

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db   1051 CGCACGCTCCTTTCCGGCTCGTCGTTTGCCGTTGAGCACGAGGAGGAGGTCAAGGTCAAG 1110

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1111 GACGACGAGGGCATTCTTCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTTCCTC 1170

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1171 GGCCCGCTCGTGGAGGACATGATGCACGCCTACTCGACTCTCTCGCTCGAGAAC---AAC 1227

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1228 ACGACGACCGACAACCCGCTCCTCGACGTCGAGAACAAGCAGACCGCGCACGGCGGCAAC 1287

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1288 TTCCAGGCGTCGGCTGTCTCGATTTCGATGGAGAAGACCAGGCTCGCACTCGCCCTCATC 1347

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1348 GGCAAGCTCAACTTCACGCAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCCTG 1407

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1408 CCTTCGTGCCTCGCTGCCGAGGACCCGTCGCTCAACTATCACGGCAAGGGCTTGGACATT 1467

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1468 CACATCGCTGCTTACGCTTCGGAGCTCGGCCACCTTGCCAACCCGGTCACTACCTTCGTC 1527

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1528 CAGCCCGCAGAGATGGGCAACCAGGCCGTCAACTCGCTCGCTCTCATCTCCGCGCGCCGC 1587

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1588 ACTGCCGAGGCCAACGACGTCCTTTCTCTCCTTCTCGCCTCGCACCTGTACTGCACGCTC 1647

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1648 CAGGCCGTCGACCTCCGCGCGATGGAGCTCGACTTCAAGAAGCAGTTCGACCCGCTTCTC 1707

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1708 CCGACTCTCCTCCAGCAGCACCTCGGCACTGGCCTCGACGTC------------AACGCA 1755

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1756 CTTGCGCTCGAGGTCAAGAAGGCGCTCAACAAGCGTCTCGAGCAGACGACGACGTACGAC 1815

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1816 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTACGCGACCGGCACCGTCGTCGAGCTCCTC 1875

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1876 TCGTCCTCGCCCTCTGCCA--ACGTCACCCTTACTGCCGTCAACGCGTGGAAGGTTGCCT 1933

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1934 CGGCCGAGAAGGCCATCTCGCTCACGCGCGAGGTGCGCAACCGCTTCTGGCAGACGCCGT 1993

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1994 CTTCGCAGGCGCCGGCGCACGCATACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2053

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2054 TGCGCGAGGAGCTCGGCGTGCAGGCGCGCCGCGGCGACGTGTTTGTCGGCGTGCAGCAGG 2113

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2114 AGACGATCGGGAGCAACGTCTCGCGCATCTACGAGGCCATCAAGGACGGCCGCATCAACC 2173

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAGNNNCNCNNNCNANNCTCGCNTNNNNNCCNNNCNC 2220
Db   2174 ACGTCCTCGTCAAGATGCTCGCGTA-AGGCCCGAGCAAGCCTCGCCTAGACGCCCGCCTC 2232
```

```
Qy       2221 NNNCCNNNNNNNNCTNTTNGNNNTCGNNTCNTGNCNNNNNCGGANNTNNCNNCNNNNNN 2280
Db       2233 ACCCCAAGACCAGCTTTTCGACGTCGTGTCGTGCCAAGAACGGACTTTCCTCCATACACA 2292

Qy       2281 TNNNNCNTNNCTNNCTCNCNNNNANCNNGTCNNTNNNNNCTNNNGNNTNNNNNNNCNNNC 2340
Db       2293 TGTCGCCTTACTCTCTCGCCGTCATCACGTCTCTCAGTTCTTTCGTATCCCGCGTCTCTC 2352

Qy       2341 NGTNNNCANNNACNCNTNNNNNNANNCNGGNANNGANTNNANGNNTNCGNGNNCNNNNNN 2400
Db       2353 GGTCGTCA-GTACACGTGTATAGAGCCTGGAATGGATTGCAAGTCTTCGAGTTCAAAAAA 2411

Qy       2401 NNNANAAA 2408
Db       2412 AAAAAAAA 2419
```

RESULT 5
US-09-765-873A-9
; Sequence 9, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-9

```
  Query Match             65.7%;  Score 1626.4;  DB 9;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476
```

```
Qy     1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db     1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy     1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db     1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy     1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db     1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy     1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db     1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy     1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db     1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy     1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db     1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy     1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db     1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy     1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db     1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy     1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db     1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy     2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db     2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy     2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db     2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy     2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db     2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 6
US-10-138-970A-23
; Sequence 23, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 23
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: mutant from Rhodotorula glutinis
US-10-138-970A-23

Query Match 65.7%; Score 1626.4; DB 15; Length 2151;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

```
Qy        1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db        1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy       61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db       61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy      121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db      101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy      181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db      161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy      241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db      218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy      301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db      277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy      361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db      337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy      421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db      397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy      481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db      457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy      541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db      517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy      601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db      577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy      661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db      637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy      721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db      697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy      781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db      757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy      841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db      817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy      901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db      877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy      961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db      937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy     1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
```

```
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 7
US-10-188-523B-9
; Sequence 9, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis mutant
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2151)
;   OTHER INFORMATION:
US-10-188-523B-9

```
  Query Match             65.7%;  Score 1626.4;  DB 15;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
```

```
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 8
US-10-374-366-3
; Sequence 3, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodosporidium glutinis
US-10-374-366-3

```
  Query Match             65.7%;  Score 1626.4;  DB 16;  Length 2151;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches 1829;  Conservative    0;  Mismatches  322;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
```

```
Db     101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy     181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240

Db     161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy     241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300

Db     218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy     301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360

Db     277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy     361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420

Db     337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy     421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480

Db     397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy     481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540

Db     457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy     541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600

Db     517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy     601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660

Db     577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy     661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720

Db     637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy     721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780

Db     697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy     781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840

Db     757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy     841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900

Db     817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy     901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960

Db     877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy     961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020

Db     937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy    1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080

Db     997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy    1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140

Db    1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy    1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200

Db    1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy    1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260

Db    1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy    1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320

Db    1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy    1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380

Db    1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy    1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440

Db    1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy    1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500

Db    1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy    1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560

Db    1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy    1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620

Db    1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy    1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680

Db    1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy    1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740

Db    1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy    1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800

Db    1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy    1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860

Db    1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy    1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920

Db    1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy    1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980

Db    1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy    1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040

Db    1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy    2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100

Db    2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy    2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160

Db    2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy    2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186

Db    2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 9
US-10-374-366-9
; Sequence 9, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song

```
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-9

  Query Match             65.6%;  Score 1624.8;  DB 16;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1828;  Conservative    0;  Mismatches  323;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGGCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885
```

```
Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 10
US-10-374-366-13
; Sequence 13, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodosporidium glutinis
US-10-374-366-13

```
  Query Match             65.6%;  Score 1624.8;  DB 16;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1828;  Conservative    0;  Mismatches  323;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCGCGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476
```

```
Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

```
RESULT 11
US-10-374-366-23
; Sequence 23, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-23

  Query Match             65.6%;  Score 1624.8;  DB 16;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1828;  Conservative    0;  Mismatches  323;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT--------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAGG 1056
```

```
Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 12
US-09-765-873A-7
; Sequence 7, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-7

Query Match 65.6%; Score 1623.2; DB 9; Length 2151;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1827; Conservative 0; Mismatches 324; Indels 35; Gaps 5;

```
Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576
```

```
Qy    601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db    577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy    661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db    637 GGCGACCTGTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy    721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db    697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy    781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db    757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy    841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db    817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTC 876

Qy    901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db    877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy    961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db    937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy   1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db    997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy   1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db   1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy   1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db   1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy   1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db   1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy   1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db   1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy   1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db   1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy   1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db   1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 13
US-10-188-523B-7
; Sequence 7, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2151)
;   OTHER INFORMATION:
US-10-188-523B-7

```
  Query Match             65.6%;  Score 1623.2;  DB 15;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1827;  Conservative    0;  Mismatches  324;  Indels   35;  Gaps    5;

Qy      1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db      1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy     61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
```

```
Db          61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy         121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db         101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy         181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db         161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy         241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db         218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy         301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db         277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy         361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db         337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy         421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db         397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy         481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db         457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy         541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db         517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy         601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db         577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy         661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db         637 GGCGACCTGTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy         721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db         697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy         781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db         757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy         841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db         817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTC 876

Qy         901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db         877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy         961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db         937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy        1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db         997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy        1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db        1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy        1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db        1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy        1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db        1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy        1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db        1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy        1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db        1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy        1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db        1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy        1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db        1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy        1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db        1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy        1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db        1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy        1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db        1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy        1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db        1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy        1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db        1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy        1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db        1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy        1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db        1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy        1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db        1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy        1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db        1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy        2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db        2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy        2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db        2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy        2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db        2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 14
US-10-374-366-1
; Sequence 1, Application US/10374366

```
; Publication No. US20040014085A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 2151
;   TYPE: DNA
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-1

  Query Match             65.6%;  Score 1623.2;  DB 16;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1827;  Conservative    0;  Mismatches  324;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGACGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCAGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTGTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAAGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCTCACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416

Qy       1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db       1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy       1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db       1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy       1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db       1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy       1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db       1597 CAAGCCATCGACTTGCGCGCGATCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy       1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db       1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy       1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db       1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy       1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db       1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836
```

```
Qy       1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db       1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy       1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db       1886 CCGCCGAGTCGGCCATCTCGCTCACCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy       1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db       1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy       2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db       2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy       2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db       2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy       2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db       2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

RESULT 15
US-10-374-366-5
; Sequence 5, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodosporidium glutinis
US-10-374-366-5

```
  Query Match             65.6%;  Score 1623.2;  DB 16;  Length 2151;
  Best Local Similarity   83.6%;  Pred. No. 0;
  Matches 1827;  Conservative    0;  Mismatches  324;  Indels   35;  Gaps    5;

Qy          1 ATGGCCCCCTCCNTCGACTCGATCGCGACCTCGNTCGCCAACGGCNTCNCNAACGGNTNG 60
Db          1 ATGGCACCCTCGCTCGACTCGATCTCGCACTCGTTCGCAAACGGCGTCGCATCCGCAAAG 60

Qy         61 CACGCCGNNCCGNCNNNGNCGNCNACGGGCGCCACGTCCACNCTCNGNCGGCCGNCGCNG 120
Db         61 CAGGCTGT---------------CAATGGCGCCTCGACCAACCTC-----GCAGTCGCAG 100

Qy        121 GCTCGCTCCTCCCGACCACCCAGNNGACGCAGCTCGACATCGTNGAGNAGATCCTCGCCG 180
Db        101 GCTCGCACCTGCCCACAACCCAGGTCACGCAGGTCGACATCGTCGAGAAGATGCTCGCCG 160

Qy        181 ACCCCACCGNNACGNACGNNNTCGAACTCGACGGGTACACCCTCACCCTCGGNGACGTCG 240
Db        161 CGCCGACCGACTCGA---CGCTCGAACTCGACGGCTACTCGCTCAACCTCGGAGACGTCG 217

Qy        241 TCGGCGCCGCNCGCAAGGGCCGCNCNGTCCGCGTCNCAGACAGNCGACGAGATCCGCGCA 300
Db        218 TCTCGGCCGCGAGGAAGGGCAGGCCTGTCCGCGTCAAGGACAG-CGACGAGATCCGCTCA 276

Qy        301 AAGATCGACAANAGNGTCGAGTTCCTCCGNNCNCAGCTCNACAACAGNGTCTACGGNGTC 360
Db        277 AAGATTGACAAATCGGTCGAGTTCTTGCGCTCGCAACTCTCCATGAGCGTCTACGGCGTC 336

Qy        361 ACGACTGGTTTCGGCGGCTCGGCCGACACCCGGACTGAGGATGCNATCTCGCTCCAGAAG 420
Db        337 ACGACTGGATTTGGCGGATCCGCAGACACCCGCACCGAGGGCGCCATCTCGCTCCAGAAG 396

Qy        421 GCNCTCCTCGAGCACCAGCTCTGCGGTGTNCTCCCNACGTCGNTCGANTCCTTCNGCCTC 480
Db        397 GCTCTCCTCGAGCACCTGCTCTGCGGTGTTCTCCCTTCGTCGTTCGACTCGTTCCGCCTC 456

Qy        481 GGNCGCGGCCTCGAGAACTCGCTTCCGCTCGAGGTCGTCCGCGGCGCCATGACCATCCGC 540
Db        457 GGCCGCGGTCTCGAGAACTCGCTTCCCCTCGAGGTTGTTCGCGGCGCCATGACAATCCGC 516

Qy        541 GTCAACTCGCTCACNCGCGGCCACTCGGCNGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 600
Db        517 GTCAACAGCTTGACCCGCGGCCACTCGGCTGTCCGCCTCGTCGTCCTCGAGGCGCTCACC 576

Qy        601 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCGGCGTCG 660
Db        577 AACTTCCTCAACCACGGCATCACCCCCATCGTCCCCCTCCGCGGCACCATCTCTGCGTCG 636

Qy        661 GGCGACCTCTCCCCNCTCTCNTACATCGCCGCCGCCATCACCGGTCACCCGGACNNCAAG 720
Db        637 GGCGACCTCTCTCCTCTCTCCTACATTGCAGCGGCCATCAGCGGTCACCCGGACAGCAAG 696

Qy        721 GTNCACGTNNTNCACGAGGGCANNGAGAAGATCATGTNCGCCCGCGAGGCGATCGCGCTC 780
Db        697 GTGCACGTCGTCCACGAGGGCAAGGAGAAGATCCTGTACGCCCGCGAGGCGATGGCGCTC 756

Qy        781 TTNGGTCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACG 840
Db        757 TTCAACCTCGAGCCCGTCGTCCTCGGCCCGAAGGAGGGTCTCGGTCTCGTCAACGGCACC 816

Qy        841 GCCGTCTCCGCCTCGATGGCGACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 900
Db        817 GCCGTCTCAGCATCGATGGCCACCCTCGCTCTGCACGACGCACACATGCTCTCGCTCCTC 876

Qy        901 TCGCAGGCGCTCACGGCTCTNACGGTCGAGGCCATGGTCGGCCACGCCGGCTCGTTCCAC 960
Db        877 TCGCAGTCGCTCACGGCCATGACGGTCGAAGCGATGGTCGGCCACGCCGGCTCGTTCCAC 936

Qy        961 CCNTTCCTCCACGACGTCACGCGCCCTCACCCGACCCAGATCGAGGTCGCGCGCAACATC 1020
Db        937 CCCTTCCTTCACGACGTCACGCGCCCTCACCCGACGCAGATCGAAGTCGCGGGAAACATC 996

Qy       1021 CGCACGCTCCTCGAGGGCAGCNNGTTTGCCGTCCACCACGAGGAGGAGGTCAAGGTCAAG 1080
Db        997 CGCAAGCTCCTCGAGGGAAGCCGCTTTGCTGTCCACCATGAGGAGGAGGTCAAGGTCAAG 1056

Qy       1081 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCGCTCCGCACGTCGCCTCAGTGGCTC 1140
Db       1057 GACGACGAGGGCATTCTCCGCCAGGACCGCTACCCCTTGCGCACGTCTCCTCAGTGGCTC 1116

Qy       1141 GGCCCGCTCGTCAGCGACATGATTCACGCCCACGCNGTCCTCTCGCTCGAGGCCGGNCAG 1200
Db       1117 GGCCCGCTCGTCAGCGACCTCATTCACGCCCACGCCGTCCTCACCATCGAGGCCGGCCAG 1176

Qy       1201 TCGACGACCGACAACCCGCTCATCGACGTCGAGAACAAGANGACCCACCACGGCGGCAAC 1260
Db       1177 TCGACGACCGACAACCCTCTCATCGACGTCGAGAACAAGACTTCGCACCACGGCGGCAAT 1236

Qy       1261 TTCCAGGCGNCCGCTGTCGCNAACACGATGGAGAAGACTCGCCTCGCNCTCGCCCTGATC 1320
Db       1237 TTCCAGGCTGCCGCTGTGGCCAACACCATGGAGAAGACTCGCCTCGGGCTCGCCCAGATC 1296

Qy       1321 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTN 1380
Db       1297 GGCAAGCTCAACTTCACGCAGCTCACCGAGATGCTCAACGCCGGCATGAACCGCGGCCTC 1356

Qy       1381 CCNTCCTGCCTCGCTGCCGAGGACCCNTCGCTCTCCTATCACTGCAAGGGCCTCGACATT 1440
Db       1357 CCCTCCTGCCTCGCGGCCGAAGACCCCTCGCTCTCCTACCACTGCAAGGGCCTCGACATC 1416
```

```
Qy   1441 GCCGCNGCNGCNTACACTTCGGAGCTCGGNCACCTTGCCAACCCGGTNACGACCCACGTC 1500
Db   1417 GCCGCTGCGGCGTACACCTCGGAGTTGGGACACCTCGCCAACCCTGTGACGACGCATGTC 1476

Qy   1501 CAGCCGGCNGAGATGGGCAACCAGGCCGTCAACTCGCTCGCNCTCATCTCGGCNCGCCGC 1560
Db   1477 CAGCCGGCTGAGATGGCGAACCAGGCGGTCAACTCGCTTGCGCTCATCTCGGCTCGTCGC 1536

Qy   1561 ACNGCCGAGGCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTNCTC 1620
Db   1537 ACGACCGAGTCCAACGACGTCCTTTCTCTCCTCCTCGCCACCCACCTCTACTGCGTTCTC 1596

Qy   1621 CAGGCCGTCGACCTCCGCGCGATGGAGTTCGAGTTCAAGAAGCAGTTCGACCCGNTNNTC 1680
Db   1597 CAAGCCATCGACTTGCGCGCGACCGAGTTCGAGTTCAAGAAGCAGTTCGGCCCAGCCATC 1656

Qy   1681 NCNNCGCTCNTCNAGCAGCACTTTGGCNCNGCCCTCGACGGCNNNNNNNNNNNNNACGAA 1740
Db   1657 GTCTCGCTCATCGACCAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTGCGCGACGAG 1716

Qy   1741 CTCGNGGACAAGGTCAACAAGNCGCTCNACAAGCGNCTCGAGCAGACCAACTCGTACGAC 1800
Db   1717 CTCGTCGAGAAGGTGAACAAGACGCTCGCCAAGCGCCTCGAGCAGACCAACTCGTACGAC 1776

Qy   1801 CTCGAGCCGCGCTGGCACGACGCCTTCTCGTTCGCGACCGGCACCGTCGTCGAGNNNNNN 1860
Db   1777 CTCGTCCCGCGCTGGCACGACGCCTTCTCCTTCGCCGCCGGCACCGTCGTCGAGGTCCTC 1836

Qy   1861 NNGTCCTCGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAACGCCTGGAAGGTCGCCT 1920
Db   1837 TCGTCGACGTC-----------GCTCTCGCTCGCCGCCGTCAACGCCTGGAAGGTCGCCG 1885

Qy   1921 CCGCCGAGAAGGCCATCTCGCTCACGCGCNANGTCCGCGACNCCTTCTGGNCGGCNCCGT 1980
Db   1886 CCGCCGAGTCGGCCATCTCGCTCACCCGCCAAGTCCGCGAGACCTTCTGGTCCGCCGCGT 1945

Qy   1981 CGTCGTCGTCGCCCGCGCTCNCGTACCTCTCGCCGCGCACGCGCGTCCTGTACTCGTTCG 2040
Db   1946 CGACCTCGTCGCCCGCGCTCTCGTACCTCTCGCCGCGCACTCAGATCCTCTACGCCTTCG 2005

Qy   2041 TCCGCGAGGAGCTCGGCGTCAAGGCCCGCCGCGGCGACGTCTTCCTCGGCAAGCAGGAGG 2100
Db   2006 TCCGCGAGGAGCTTGGCGTCAAGGCCCGCCGCGGAGACGTCTTCCTCGGCAAGCAAGAGG 2065

Qy   2101 TGACGATCGGCACCAACGTCTCCCGCATCTACGAGGCCATCAAGNNCGGCNGCATCAACC 2160
Db   2066 TGACGATCGGCTCGAACGTCTCCAAGATCTACGAGGCCATCAAGTCGGGCAGGATCAACA 2125

Qy   2161 ACGTCCTCGTCAAGATGCTCGCNTAG 2186
Db   2126 ACGTCCTCCTCAAGATGCTCGCTTAG 2151
```

Search completed: September 12, 2004, 12:22:21
Job time : 1619.68 secs

OM protein - protein search, using sw model

```
Run on:          September 9, 2004, 10:01:03 ; Search time 23.5975 Seconds
                                          (without alignments)
                                          1588.322 Million cell updates/sec

Title:           US-09-939-408A-21
Perfect score:   3365
Sequence:        1 MAPSLDSIATSXANGXXNGX..........RIYEAIKSGRINXVLVKMLA 726

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3223	95.8	726	4	US-09-624-693A-21	Sequence 21, Appl
2	3021	89.8	716	4	US-09-624-693A-19	Sequence 19, Appl
3	3021	89.8	716	4	US-09-627-216A-8	Sequence 8, Appli
4	3021	89.8	716	4	US-09-765-873A-8	Sequence 8, Appli
5	3019	89.7	716	4	US-09-627-216A-10	Sequence 10, Appl
6	3019	89.7	716	4	US-09-765-873A-10	Sequence 10, Appl
7	3014	89.6	716	4	US-09-765-873A-38	Sequence 38, Appl
8	3012	89.5	716	4	US-09-765-873A-35	Sequence 35, Appl
9	3010	89.5	716	4	US-09-765-873A-36	Sequence 36, Appl
10	3010	89.5	716	4	US-09-765-873A-37	Sequence 37, Appl
11	3006	89.3	716	4	US-09-765-873A-34	Sequence 34, Appl
12	2978	88.5	716	4	US-09-765-873A-33	Sequence 33, Appl
13	2971	88.3	686	4	US-09-765-873A-32	Sequence 32, Appl
14	2922.5	86.8	713	4	US-09-624-693A-17	Sequence 17, Appl
15	2770	82.3	720	4	US-09-624-693A-13	Sequence 13, Appl
16	1294.5	38.5	740	4	US-09-624-693A-15	Sequence 15, Appl
17	670	19.9	424	4	US-09-615-192A-328	Sequence 328, App
18	495.5	14.7	515	4	US-09-328-352-5090	Sequence 5090, Ap
19	490	14.6	508	4	US-09-489-039A-13575	Sequence 13575, A
20	480.5	14.3	515	4	US-09-252-991A-33046	Sequence 33046, A
21	474.5	14.1	525	4	US-09-540-236-2855	Sequence 2855, Ap
22	459	13.6	412	4	US-09-615-192A-330	Sequence 330, App
23	365	10.8	520	4	US-09-252-991A-33049	Sequence 33049, A
24	233	6.9	226	4	US-09-615-192A-327	Sequence 327, App
25	223.5	6.6	97	4	US-09-615-192A-329	Sequence 329, App
26	115	3.4	1024	4	US-09-562-737-50	Sequence 50, Appl
27	111	3.3	1024	4	US-09-562-737-41	Sequence 41, Appl
28	108	3.2	1024	4	US-09-562-737-44	Sequence 44, Appl
29	106	3.2	505	4	US-09-252-991A-23925	Sequence 23925, A
30	104	3.1	744	1	US-08-764-100-25	Sequence 25, Appl
31	102.5	3.0	1011	4	US-09-252-991A-22505	Sequence 22505, A
32	100.5	3.0	555	4	US-09-543-681A-7878	Sequence 7878, Ap
33	100	3.0	1589	4	US-09-543-681A-4998	Sequence 4998, Ap
34	99.5	3.0	1024	4	US-09-562-737-45	Sequence 45, Appl
35	96.5	2.9	404	4	US-09-328-352-6480	Sequence 6480, Ap
36	96	2.9	875	4	US-09-328-352-4884	Sequence 4884, Ap
37	95.5	2.8	836	4	US-09-252-991A-23513	Sequence 23513, A
38	94.5	2.8	607	3	US-08-969-683A-59	Sequence 59, Appl
39	94	2.8	555	1	US-08-440-377A-6	Sequence 6, Appli
40	94	2.8	555	2	US-08-687-852-6	Sequence 6, Appli
41	94	2.8	764	4	US-09-252-991A-31816	Sequence 31816, A
42	93.5	2.8	1024	4	US-09-562-737-47	Sequence 47, Appl
43	93.5	2.8	1024	4	US-09-562-737-49	Sequence 49, Appl
44	93.5	2.8	1495	4	US-08-522-726B-1	Sequence 1, Appli
45	93.5	2.8	1495	4	US-09-337-384-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-624-693A-21
; Sequence 21, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSC
;  CURRENT APPLICATION NUMBER: US/09/624,693A
;  CURRENT FILING DATE:  2000-07-24
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 726
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (12, 16, 17, 20, 24, 25, 28 - 36, 38, 42, 47, 48, 56, 57, 62, 66,
;   OTHER INFORMATION: "Xaa" means any amino acid; "Xaa" means no consensus at that
;   OTHER INFORMATION: position
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus of
;   OTHER INFORMATION: SEQ ID NOs: 13, 17, and 19
US-09-624-693A-21

  Query Match             95.8%;  Score 3223;  DB 4;  Length 726;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  726;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        721 LVKMLA 726
              ||||||
Db        721 LVKMLA 726
```

RESULT 2
US-09-624-693A-19
; Sequence 19, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula toruloides
US-09-624-693A-19

```
  Query Match             89.8%;  Score 3021;  DB 4;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||:  | |||   :    |   ||          | |||  ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||    |||||| |||||  ||||| |||  ||||||  |||||||||  ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||:| |||  ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || |||  |||:  ||||:||  |||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||||:||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||| |||||||||||||| |:||||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||||| |     |: ||||         ||  || | | ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||       :|||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:|||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 3
US-09-627-216A-8
; Sequence 8, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-09-627-216A-8

```
  Query Match             89.8%;  Score 3021;  DB 4;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||:  | |||   :    |   ||          | |||  ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
```

```
Db          55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy         121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Db         114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy         181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Db         174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy         241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Db         234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy         301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Db         294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy         361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Db         354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy         421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Db         414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy         481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Db         474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy         541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Db         534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy         601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Db         594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy         661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Db         651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy         721 LVKMLA 726

Db         711 LLKMLA 716
```

RESULT 4
US-09-765-873A-8
; Sequence 8, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-09-765-873A-8

```
  Query Match             89.8%;  Score 3021;  DB 4;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;
```

```
Qy           1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Db           1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy          61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Db          55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy         121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Db         114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy         181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Db         174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy         241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Db         234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy         301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Db         294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy         361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Db         354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy         421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Db         414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy         481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Db         474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy         541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Db         534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy         601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Db         594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy         661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Db         651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy         721 LVKMLA 726

Db         711 LLKMLA 716
```

RESULT 5
US-09-627-216A-10
; Sequence 10, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10

```
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: mutant from
;   OTHER INFORMATION: Rhodtorula glutinis
US-09-627-216A-10

  Query Match             89.7%;  Score 3019;  DB 4;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              |||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||||||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||   |      ||  || | | |||||||||||
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

RESULT 6
US-09-765-873A-10
; Sequence 10, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-10

  Query Match             89.7%;  Score 3019;  DB 4;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 0;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              |||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||||||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||   |      ||  || | | |||||||||||
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

RESULT 7
US-09-765-873A-38
```

```
; Sequence 38, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 38
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

  Query Match             89.6%;  Score 3014;  DB 4;  Length 716;
  Best Local Similarity   83.9%;  Pred. No. 0;
  Matches  609;  Conservative   24;  Mismatches   83;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716

RESULT 8
US-09-765-873A-35
; Sequence 35, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: X= Asp, Asn, Glu, Gln
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-35

  Query Match             89.5%;  Score 3012;  DB 4;  Length 716;
  Best Local Similarity   83.9%;  Pred. No. 0;
  Matches  609;  Conservative   24;  Mismatches   83;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHXITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
```

```
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716
```

RESULT 9
US-09-765-873A-36
; Sequence 36, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (181)..(181)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (235)..(235)
;   OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-36

```
  Query Match             89.5%;  Score 3010;  DB 4;  Length 716;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches  608;  Conservative   24;  Mismatches   84;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHXAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HXVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716
```

RESULT 10
US-09-765-873A-37
; Sequence 37, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (149)..(149)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (202)..(202)
;   OTHER INFORMATION: X= Val, Met, Leu, Cys
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-37

```
  Query Match             89.5%;  Score 3010;  DB 4;  Length 716;
```

```
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches  608;  Conservative   24;  Mismatches   84;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDXFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPXVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716

RESULT 11
US-09-765-873A-34
; Sequence 34, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 34
;  LENGTH: 716
;  TYPE: PRT
;  ORGANISM: mutant from Rhodotorula glutinis
;  FEATURE:
;  NAME/KEY: UNSURE
;  LOCATION: (126)..(126)
;  OTHER INFORMATION: X= Gly, Ala, Ser, Thr
;  NAME/KEY: UNSURE
;  LOCATION: (138)..(138)
;  OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
;  NAME/KEY: UNSURE
;  LOCATION: (540)..(540)
;  OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-34

  Query Match             89.3%;  Score 3006;  DB 4;  Length 716;
  Best Local Similarity   83.7%;  Pred. No. 0;
  Matches  608;  Conservative   24;  Mismatches   84;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEXAISLQKALLEHXLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716

RESULT 12
```

```
US-09-765-873A-33
; Sequence 33, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (126)..(126)
;   OTHER INFORMATION: X= Gly, Ser, Ala, Thr
;   NAME/KEY: UNSURE
;   LOCATION: (138)..(138)
;   OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
;   NAME/KEY: UNSURE
;   LOCATION: (149)..(149)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (181)..(181)
;   OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
;   NAME/KEY: UNSURE
;   LOCATION: (198)..(198)
;   OTHER INFORMATION: X= Asp, Asn, Glu, Gln
;   NAME/KEY: UNSURE
;   LOCATION: (202)..(202)
;   OTHER INFORMATION: X= Val, Met, Leu, Cys
;   NAME/KEY: UNSURE
;   LOCATION: (235)..(235)
;   OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-09-765-873A-33

  Query Match             88.5%;  Score 2978;  DB 4;  Length 716;
  Best Local Similarity   82.9%;  Pred. No. 0;
  Matches  602;  Conservative   25;  Mismatches   89;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||   :   |    ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||    |||||| ||||| |||||| ||| |||||| ||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||| |||||||||| ||||||:| | ||||||||||||||||||||||
Db        114 TGFGGSADTRTEXAISLQKALLEHXLCGVLPSSFDXFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              |||||| ||||||||||||||||| | ||||||||||||||||||||||||:||||||
Db        174 NSLTRGHXAVRLVVLEALTNFLNHXITPXVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              | ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||
Db        234 HXVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||| ||| ||||| |||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:||||||::||||||||||||||||| :|||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || |||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||| ||| |||||||||| |:||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||| |     |: ||||    |         ||  || | | ||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               ||||||| | ||||| |||        :||||||||| || ||||| ||  || | :
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| |||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716


RESULT 13
US-09-765-873A-32
; Sequence 32, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
;   LENGTH: 686
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-32

  Query Match             88.3%;  Score 2971;  DB 4;  Length 686;
  Best Local Similarity   86.2%;  Pred. No. 0;
  Matches  595;  Conservative   23;  Mismatches   68;  Indels    4;  Gaps    2;

Qy         37 AXAGSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDE 96
              | ||| ||||  ||:||||  || | ||    |||||| ||||| |||||| ||| ||||
Db          1 AVAGSHLPTTQVTQVDIVEKMLAAP-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDE 59

Qy         97 IRXKIDKSVEFLRXQLXNSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDS 156
              || ||||||||| ||  ||||||||||||||||||||||||||||||||||||:| ||
Db         60 IRSKIDKSVEFLRSQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDS 119

Qy        157 FXLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTI 216
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 FRLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTI 179

Qy        217 SASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLV 276
              |||||||||||||||:||||||||| ||| |||: ||||:||| ||||||||||||||
```

```
Db        180 SASGDLSPLSYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLV 239

Qy        277 NGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVA 336

Db        240 NGTAVSASMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVA 299

Qy        337 RNIRTLLEGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLE 396

Db        300 GNIRKLLEGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIE 359

Qy        397 AGQSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMN 456

Db        360 AGQSTTDNPLIDVENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMN 419

Qy        457 RGLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALIS 516

Db        420 RGLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALIS 479

Qy        517 ARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXX 576

Db        480 ARRTTESNDVLSLLLATHLYCVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNL 539

Qy        577 XXELXXKVXKXLXKRLEQTNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWK 636

Db        540 RDELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWK 596

Qy        637 VAXAEKAISLTRXVRXXFWXAPSSSSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGK 696

Db        597 VAAAESAISLTRQVRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDVFLGK 656

Qy        697 QEVTIGSNVSRIYEAIKSGRINXVLVKMLA 726

Db        657 QEVTIGSNVSKIYEAIKSGRINNVLLKMLA 686
```

RESULT 14
US-09-624-693A-17
; Sequence 17, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Rhodotorula mucilaginosa
US-09-624-693A-17

```
  Query Match             86.8%;  Score 2922.5;  DB 4;  Length 713;
  Best Local Similarity   82.0%;  Pred. No. 0;
  Matches  595;  Conservative   33;  Mismatches   85;  Indels   13;  Gaps    5;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Db          1 MAPSVDSIATSVANSLSNGLHAAAAAN-GGDVHKKTAGAGSLLPTTETTQLDIVERILAD 59

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Db         60 AGATDQIKLDGYTLTLGDVVGAARRGRSVKVADSPHIREKIDASVEFLRTQLDNSVYGVT 119

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Db        120 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSMDGFALGRGLENSLPLEVVRGAMTIRV 179

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Db        180 NSLTRGHSAVRIVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAASITGHPDSKV 239

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Db        240 HV--DG--KIMSAQEAIALKGLQPVVLGPKEGLGLVNGTAVSASMATLALTDAHVLSLLA 295

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Db        296 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSKYAVHHETEVKVKD 355

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Db        356 DEGILRQDRYPLRCSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDLENKMTHHGGAF 415

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Db        416 MASSVGNTMEKTRLAVALMGKVSFTQLTEMLNAGMNRALPSCLAAEDPSLSYHCKGLDIA 475

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Db        476 AAAYTSELGHLANPVSTHVQPAEMGNQAINSLALISARRTAEANDVLSLLLATHLYCVLQ 535

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Db        536 AVDLRAMEFEHTKAFEPMVTELLKQHFGALAT-----AEVEDKVRKSIYKRLQQNNSYDL 590

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Db        591 EQRWHDTFSVATGAVVEALAG---QEVSLASLNAWKVACAEKAIALTRSVRDSFWAAPSS 647

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Db        648 SSPALKYLSPRTRVLYSFVREEVGVKARRGDVYLGKQEVTIGTNVSRIYEAIKSGCIAPV 707

Qy        721 LVKMLA 726

Db        708 LVKMMA 713
```

RESULT 15
US-09-624-693A-13
; Sequence 13, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Rhodotorula graminis
US-09-624-693A-13

```
  Query Match             82.3%;  Score 2770;  DB 4;  Length 720;
  Best Local Similarity   77.5%;  Pred. No. 4.6e-309;
  Matches  563;  Conservative   40;  Mismatches  117;  Indels    6;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Db          1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSALRRTPGLDGHAAHQSQLEIVQELLSD 60

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Db         61 P-TDDVVELSGYSLTVRDVVGAARKGRRVRVQNDDEIRARVDKSVDFLKAQLQNSVYGVT 119
```

```
Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Db        120 TGFGGSADTRTEDAVSLQKALIEHQLCGVTPTSVSSFSVGRGLENTLPLEVVRGAMVIRV 179

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Db        180 NSLTRGHSAVRLVVLEALTNFLNHRITPIVPLRGSISASGDLSPLSYIAGAITGHPDVKV 239

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Db        240 HVLHEGTEKIMFAREAISLFGLEAVVLGPKEGLGLVNGTAVSASMATLSLHDSHMLSLLS 299

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Db        300 QALTALTVEAMVGQQGSFAPFIHDVCRPHPGQVEVARNIRTLLSGSSFAVEHEEEVKVKD 359

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Db        360 DEGILRQDRYPLRTSPQFLGPLVEDMMHAYSTLSLE-NNTTTDNPLLDVENKQTAHGGNF 418

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Db        419 QASAVSISMEKTRLALALIGKLNFTQCTELLNAAMNRGLPSCLAAEDPSLNYHGKGLDIH 478

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Db        479 IAAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLLASHLYCTLQ 538

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Db        539 AVDLRAMELDFKKQFDPLLPTLLQQHLGTGLDVNA----LALEVKKALNKRLEQTTTYDL 594

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Db        595 EPRWHDAFSYATGTVVELLSSSPSANVTLTAVNAWKVASAEKAISLTREVRNRFWQTPSS 654

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Db        655 QAPAHAYLSPRTRVLYSFVREELGVQARRGDVFVGVQQETIGSNVSRIYEAIKDGRINHV 714

Qy        721 LVKMLA 726

Db        715 LVKMLA 720
```

Search completed: September 9, 2004, 10:09:41
Job time : 26.5975 secs

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:07:54 ; Search time 93.3859 Seconds
(without alignments)
2493.103 Million cell updates/sec

Title: US-09-939-408A-21
Perfect score: 3365
Sequence: 1 MAPSLDSIATSXANGXXNGX..........RIYEAIKSGRINXVLVKMLA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3223	95.8	726	9	US-09-939-408A-21	Sequence 21, Appl
2	3021	89.8	716	9	US-09-765-873A-8	Sequence 8, Appli
3	3021	89.8	716	9	US-09-939-408A-19	Sequence 19, Appl
4	3021	89.8	716	14	US-10-138-970A-4	Sequence 4, Appli
5	3021	89.8	716	14	US-10-188-523B-8	Sequence 8, Appli
6	3021	89.8	716	15	US-10-374-366-2	Sequence 2, Appli
7	3021	89.8	716	15	US-10-439-478-2	Sequence 2, Appli
8	3019	89.7	716	9	US-09-765-873A-10	Sequence 10, Appl
9	3019	89.7	716	14	US-10-138-970A-24	Sequence 24, Appl
10	3019	89.7	716	14	US-10-188-523B-10	Sequence 10, Appl
11	3019	89.7	716	15	US-10-374-366-4	Sequence 4, Appli
12	3016	89.6	716	15	US-10-374-366-24	Sequence 24, Appl
13	3015	89.6	716	15	US-10-374-366-22	Sequence 22, Appl
14	3014	89.6	716	9	US-09-765-873A-38	Sequence 38, Appl
15	3014	89.6	716	14	US-10-138-970A-22	Sequence 22, Appl
16	3014	89.6	716	14	US-10-188-523B-38	Sequence 38, Appl
17	3014	89.6	716	15	US-10-374-366-16	Sequence 16, Appl
18	3014	89.6	716	15	US-10-439-478-18	Sequence 18, Appl
19	3013	89.5	716	15	US-10-374-366-8	Sequence 8, Appli
20	3012	89.5	716	9	US-09-765-873A-35	Sequence 35, Appl
21	3012	89.5	716	14	US-10-138-970A-19	Sequence 19, Appl
22	3012	89.5	716	14	US-10-188-523B-35	Sequence 35, Appl
23	3012	89.5	716	15	US-10-374-366-10	Sequence 10, Appl
24	3012	89.5	716	15	US-10-439-478-15	Sequence 15, Appl
25	3011	89.5	716	15	US-10-374-366-14	Sequence 14, Appl
26	3010	89.5	716	9	US-09-765-873A-36	Sequence 36, Appl
27	3010	89.5	716	9	US-09-765-873A-37	Sequence 37, Appl
28	3010	89.5	716	14	US-10-138-970A-20	Sequence 20, Appl
29	3010	89.5	716	14	US-10-138-970A-21	Sequence 21, Appl
30	3010	89.5	716	14	US-10-188-523B-36	Sequence 36, Appl
31	3010	89.5	716	14	US-10-188-523B-37	Sequence 37, Appl
32	3010	89.5	716	15	US-10-439-478-16	Sequence 16, Appl
33	3010	89.5	716	15	US-10-439-478-17	Sequence 17, Appl
34	3006	89.3	716	9	US-09-765-873A-34	Sequence 34, Appl
35	3006	89.3	716	14	US-10-138-970A-18	Sequence 18, Appl
36	3006	89.3	716	14	US-10-188-523B-34	Sequence 34, Appl
37	3006	89.3	716	15	US-10-374-366-18	Sequence 18, Appl
38	3006	89.3	716	15	US-10-439-478-14	Sequence 14, Appl
39	3005	89.3	716	15	US-10-374-366-6	Sequence 6, Appli
40	3005	89.3	716	15	US-10-374-366-20	Sequence 20, Appl
41	2997	89.1	716	15	US-10-374-366-12	Sequence 12, Appl
42	2978	88.5	716	9	US-09-765-873A-33	Sequence 33, Appl
43	2978	88.5	716	14	US-10-138-970A-17	Sequence 17, Appl
44	2978	88.5	716	14	US-10-188-523B-33	Sequence 33, Appl
45	2978	88.5	716	15	US-10-439-478-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-939-408A-21
; Sequence 21, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)..(719)
; OTHER INFORMATION: "Xaa"means any amino acid;"Xaa"means no consensus at that posit
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOs: 13, 17, and 19
US-09-939-408A-21

Query Match 95.8%; Score 3223; DB 9; Length 726;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
```

```
Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        721 LVKMLA 726
              ||||||
Db        721 LVKMLA 726
```

```
RESULT 2
US-09-765-873A-8
; Sequence 8, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
US-09-765-873A-8

  Query Match             89.8%;  Score 3021;  DB 9;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.1e-298;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              |||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||||||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:|||||||| |    |: ||||   |      ||  || | | |||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

RESULT 3
US-09-939-408A-19
; Sequence 19, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
;  APPLICANT: Yoshida, Roberta
;  APPLICANT:  Kootstra, Anna
;  TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
;  TITLE OF INVENTION:  Polynucleotide Sequences and Methods of Obtaining and
;  TITLE OF INVENTION:  Using Same
;  FILE REFERENCE: 29479/500NSCA
;  CURRENT APPLICATION NUMBER: US/09/939,408A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 09/624,693
;  PRIOR FILING DATE: 2000-07-24
;  PRIOR APPLICATION NUMBER: PCT/US01/23270
;  PRIOR FILING DATE: 2001-07-24
```

```
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula toruloides
US-09-939-408A-19

  Query Match             89.8%;  Score 3021;  DB 9;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.1e-298;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||  :   |   ||          | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||    |||||| ||||| |||||| ||| ||||||| ||||||||| || ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:||||||::|||||||||||||||||||| :|||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||| ||||||||||||||| |:||||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||| |    |: ||||   |        ||  || | |  ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               ||||||| | ||||| |||       :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 4
US-10-138-970A-4
; Sequence 4, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: C. violaceum
US-10-138-970A-4

```
  Query Match             89.8%;  Score 3021;  DB 14;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.1e-298;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||  :   |   ||          | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||    |||||| ||||| |||||| ||| ||||||| ||||||||| || ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:||||||::|||||||||||||||||||| :|||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||| ||||||||||||||| |:||||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||| |    |: ||||   |        ||  || | |  ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               ||||||| | ||||| |||       :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 5
US-10-188-523B-8
; Sequence 8, Application US/10188523B

```
; Publication No. US20030170834A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US DIVCIP
;  CURRENT APPLICATION NUMBER: US/10/188,523B
;  CURRENT FILING DATE:  2002-10-17
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
US-10-188-523B-8

  Query Match             89.8%;  Score 3021;  DB 14;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.1e-298;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716

RESULT 6
US-10-374-366-2
; Sequence 2, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-2

  Query Match             89.8%;  Score 3021;  DB 15;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.1e-298;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
```

```
             ||||| ||||||::||:||||||||||||||||||||||||||||||:|||||||||| |
Db       651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy       721 LVKMLA 726
             |:||||
Db       711 LLKMLA 716


RESULT 7
US-10-439-478-2
; Sequence 2, Application US/10439478
; Publication No. US20040018600A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. du Pont de Nemours and Co.
;  APPLICANT:  Ben-Bassat, Arie
;  APPLICANT:  Qi, Wei Wei
;  APPLICANT:  Sariaslani, Sima
;  APPLICANT:  Tang, Xao-Song
;  APPLICANT:  Vanelli, Todd
;  TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
;  FILE REFERENCE: CL1912
;  CURRENT APPLICATION NUMBER: US/10/439,478
;  CURRENT FILING DATE:  2003-05-16
;  PRIOR APPLICATION NUMBER: 60/383450
;  PRIOR FILING DATE: 2002-05-23
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
US-10-439-478-2

  Query Match             89.8%;  Score 3021;  DB 15;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.1e-298;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy         1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
             ||||||||: | |||  :    |   ||             | ||| ||||  ||:|||| ||
Db         1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy        61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
             | ||   |||||| ||||| ||||| ||| |||||| ||||||||| ||   |||||
Db        55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy       121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
             |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db       114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy       181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy       241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
             || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||||
Db       234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy       301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
             |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db       294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy       361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
             ||||||||||||||||||||||||||:||||||::||||||||||||||||| :||||||
Db       354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy       421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
             ||:|| ||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db       414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy       481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
             ||||||||||||||||||||||| ||||||||||||||| |:||||||||||||||||||
Db       474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy       541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
             |:||||:||||||||  |    |: ||||   |        ||  ||  | | ||||||||||
Db       534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy       601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              ||||||| | ||||| |||      :||||||||||| || |||||| ||   || | |:
Db       594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy       661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
             ||||| |||||::||:||||||||||||||||||||||||||||||||:|||||||||| |
Db       651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy       721 LVKMLA 726
             |:||||
Db       711 LLKMLA 716


RESULT 8
US-09-765-873A-10
; Sequence 10, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
;  FILE REFERENCE: BC1009 US CIP
;  CURRENT APPLICATION NUMBER: US/09/765,873A
;  CURRENT FILING DATE:  2001-01-19
;  PRIOR APPLICATION NUMBER: US 09/627,216
;  PRIOR FILING DATE: 2000-07-27
;  PRIOR APPLICATION NUMBER: US 60/147,719
;  PRIOR FILING DATE: 1999-08-06
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-10

  Query Match             89.7%;  Score 3019;  DB 9;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 1.1e-297;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy         1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
             ||||||||: | |||  :    |   ||             | ||| ||||  ||:|||| ||
Db         1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy        61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
             | ||   |||||| ||||| ||||| ||| |||||| ||||||||| ||   |||||
Db        55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy       121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
             |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db       114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy       181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db       174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy       241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
             || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||||
Db       234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy       301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
             |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db       294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy       361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
             ||||||||||||||||||||||||||:||||||::||||||||||||||||| :||||||
Db       354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy       421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
             ||:|| ||||||| || |||||||||||||||||||||||||||||||||||||||||||
```

```
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||||||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||   |      ||  || | | |||||||||||
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 9
US-10-138-970A-24
; Sequence 24, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sariasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 716
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-138-970A-24

```
  Query Match             89.7%;  Score 3019;  DB 14;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 1.1e-297;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  || 
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              |||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||||||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||   |      ||  || | | |||||||||||
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 10
US-10-188-523B-10
; Sequence 10, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis mutant
US-10-188-523B-10

```
  Query Match             89.7%;  Score 3019;  DB 14;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 1.1e-297;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  || 
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
```

```
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716
```

```
RESULT 11
US-10-374-366-4
; Sequence 4, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-4

  Query Match             89.7%;  Score 3019;  DB 15;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 1.1e-297;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716
```

```
RESULT 12
US-10-374-366-24
; Sequence 24, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Xiao-Song
;  APPLICANT:  Milano, Joseph
;  TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
;  FILE REFERENCE: CL1794 US NA
;  CURRENT APPLICATION NUMBER: US/10/374,366
;  CURRENT FILING DATE:  2003-02-26
;  PRIOR APPLICATION NUMBER: 60/360,279
;  PRIOR FILING DATE: 2002-02-26
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 24
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodosporidium glutinis
US-10-374-366-24

  Query Match             89.6%;  Score 3016;  DB 15;  Length 716;
  Best Local Similarity   83.9%;  Pred. No. 2.3e-297;
  Matches  609;  Conservative   25;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113
```

```
Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:||||||||||||||||||||||||||||||| ||| ||||| |||||||||||:|
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVRD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:|||||||::||||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| |||||||||||||||| |:||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||    |       ||   || | | ||||||||||
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| ||||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| |||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 13
US-10-374-366-22
; Sequence 22, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodosporidium glutinis
US-10-374-366-22

```
  Query Match            89.6%;  Score 3015;  DB 15;  Length 716;
  Best Local Similarity  83.9%;  Pred. No. 2.9e-297;
  Matches  609;  Conservative   24;  Mismatches   83;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||  :   |   ||          | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||    |||||| ||||| ||||| ||| |||||| ||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:|||||||::||||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| |||||||||||||||| |:||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||    |       ||   || | | ||||||||||
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :|||||| |||| ||||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNTWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| |||||::||:|||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 14
US-09-765-873A-38
; Sequence 38, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 38
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE

```
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
;   NAME/KEY: UNSURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

  Query Match             89.6%;  Score 3014;  DB 9;  Length 716;
  Best Local Similarity   83.9%;  Pred. No. 3.7e-297;
  Matches  609;  Conservative   24;  Mismatches   83;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              |||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||| ||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXXKLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||   |      ||  ||   | |||||||||||
Db        534 AIDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:||||||||||||||||||||||||||||||:|||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

RESULT 15
US-10-138-970A-22
; Sequence 22, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
;  APPLICANT: Sariasiani, Sima
;  APPLICANT:  Tang, Xiao-Song
;  APPLICANT:  Qi, Wei Wei
;  TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
;  TITLE OF INVENTION:  hydroxycinnamic Acid
;  FILE REFERENCE: CL1777
;  CURRENT APPLICATION NUMBER: US/10/138,970A
;  CURRENT FILING DATE:  2002-07-23
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Rhodotorula glutinis
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (502)..(502)
;   OTHER INFORMATION: x=Gly, Ala, Ser,Thr, Pro
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: x=Thr, Ala, Ser, Pro, Gly
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (540)..(540)
;   OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-138-970A-22

  Query Match             89.6%;  Score 3014;  DB 14;  Length 716;
  Best Local Similarity   83.9%;  Pred. No. 3.7e-297;
  Matches  609;  Conservative   24;  Mismatches   83;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              |||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||| ||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAXNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXXKLXKRLEQTNSYDL 600
              |:|||| |||||||| |    |: ||||   |      ||  ||   | |||||||||||
Db        534 AIDLRAXEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
```

```
             ||||| ||||||::||:||||||||||||||||||||||||||||:||||||||||||  |
Db       651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy       721 LVKMLA 726
             |:||||
Db       711 LLKMLA 716

Search completed: September  9, 2004, 10:24:27
Job time : 96.3859 secs
```

GenCore version 5.1.6


```
OM protein - protein search, using sw model

Run on:          September  9, 2004, 09:50:23 ; Search time 16.5685 Seconds
                                           (without alignments)
                                           2281.620 Million cell updates/sec

Title:           US-09-939-408A-21
Perfect score:   3365
Sequence:        1 MAPSLDSIATSXANGXXNGX..........RIYEAIKSGRINXVLVKMLA 726

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3021	89.8	716	1	PALY_RHOTO	P11544 rhodosporid
2	2922.5	86.8	713	1	PALY_RHORB	P10248 rhodotorula
3	1294.5	38.5	740	1	PALY_AMAMU	O93967 amanita mus
4	1205	35.8	724	1	PALY_USTMA	Q96v77 ustilago ma
5	965	28.7	725	1	PAL1_ARATH	P35510 arabidopsis
6	958.5	28.5	718	1	PAL3_PETCR	P45729 petroselinu
7	957	28.4	754	1	PALY_PINTA	P52777 pinus taeda
8	954	28.4	725	1	PALY_MEDSA	P27990 medicago sa
9	952.5	28.3	701	1	PAL1_ORYSA	P14717 oryza sativ
10	949	28.2	725	1	PALY_TRISU	P45734 trifolium s
11	947	28.1	717	1	PAL1_PRUAV	O64963 prunus aviu
12	946.5	28.1	716	1	PAL1_PETCR	P24481 petroselinu
13	946.5	28.1	716	1	PAL2_PETCR	P45728 petroselinu
14	946	28.1	713	1	PAL1_SOYBN	P27991 glycine max
15	943.5	28.0	707	1	PAL4_ARATH	Q9ss45 arabidopsis
16	943	28.0	715	1	PAL1_TOBAC	P25872 nicotiana t
17	940	27.9	713	1	PALY_DIGLA	O23924 digitalis l
18	939.5	27.9	717	1	PAL2_ARATH	P45724 arabidopsis
19	938	27.9	712	1	PAL3_TOBAC	P45733 nicotiana t
20	933.5	27.7	704	1	PAL1_LYCES	P35511 lycopersico
21	932	27.7	723	1	PAL1_PEA	Q01861 pisum sativ
22	931.5	27.7	710	1	PAL1_RUBID	Q9m568 rubus idaeu
23	930.5	27.7	721	1	PAL5_LYCES	P26600 lycopersico
24	929.5	27.6	714	1	PALY_CAMSI	P45726 camellia si
25	929	27.6	667	1	PALY_HELAN	O04058 helianthus
26	929	27.6	712	1	PAL2_TOBAC	P35513 nicotiana t
27	928.5	27.6	712	1	PAL2_PHAVU	P19142 phaseolus v
28	927.5	27.6	708	1	PAL2_IPOBA	Q42858 ipomoea bat
29	927	27.5	708	1	PAL1_DAUCA	O23865 daucus caro
30	927	27.5	710	1	PAL2_ORYSA	P53443 oryza sativ
31	926.5	27.5	724	1	PAL2_PEA	Q04593 pisum sativ
32	925	27.5	705	1	PAL2_LITER	O49836 lithospermu
33	924.5	27.5	715	1	PALY_STYHU	P45732 stylosanthe
34	923.5	27.4	718	1	PAL2_CICAR	Q9smk9 cicer ariet
35	922	27.4	710	1	PAL1_LITER	O49835 lithospermu
36	918	27.3	590	1	PAL2_SOLTU	P31426 solanum tub
37	917.5	27.3	722	1	PALY_CITLI	Q42667 citrus limo
38	916.5	27.2	720	1	PAL1_SOLTU	P31425 solanum tub
39	910.5	27.1	715	1	PALY_POPTR	P45730 populus tri
40	910	27.0	695	1	PAL3_ARATH	P45725 arabidopsis
41	910	27.0	710	1	PAL3_PHAVU	P19143 phaseolus v
42	906.5	26.9	710	1	PAL2_POPKI	Q43052 populus kit
43	897.5	26.7	682	1	PAL1_POPKI	P45731 populus kit
44	893.5	26.6	707	1	PAL1_IPOBA	P14166 ipomoea bat
45	891	26.5	700	1	PALY_WHEAT	Q43210 triticum ae

ALIGNMENTS

```
RESULT 1
PALY_RHOTO
ID   PALY_RHOTO     STANDARD;      PRT;   716 AA.
AC   P11544;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Rhodosporidium toruloides (Yeast) (Rhodotorula gracilis).
OC   Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
OC   Microbotryomycetidae; Sporidiobolales; Rhodosporidium.
OX   NCBI_TaxID=5286;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 14;
RX   MEDLINE=92127061; PubMed=1773059;
RA   Rasmussen O.F., Orum H.;
RT   "Analysis of the gene for phenylalanine ammonia-lyase from
RT   Rhodosporidium toruloides.";
RL   DNA Seq. 1:207-211(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 0559;
RX   MEDLINE=88112870; PubMed=2828184;
RA   Anson J.G., Gilbert H.J., Oram J.D., Minton N.P.;
RT   "Complete nucleotide sequence of the Rhodosporidium toruloides gene
RT   coding for phenylalanine ammonia-lyase.";
RL   Gene 58:189-199(1987).
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X51513; CAA35886.1; -.
DR   EMBL; M18261; AAA33883.1; -.
DR   EMBL; X12702; CAA31209.1; -.
DR   PIR; A29607; A29607.
DR   PIR; A56628; A56628.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
```

```
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    211    213       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     212    212       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
FT   CONFLICT      4     37       SLDSISHSFANGVASAKQAVNGASTNLAVAGSHL ->
FT                                RPTSQSQARTC (IN REF. 2).
SQ   SEQUENCE   716 AA;  76879 MW;  0C1DF61769A4E5E6 CRC64;

  Query Match             89.8%;  Score 3021;  DB 1;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 9.8e-227;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||  :   |   ||            | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   |||||| ||||| |||||| ||| |||||| ||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:||||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || |||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||| ||||||||||||||| |:||||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:|||||||| |     |: ||||  |       ||  || | | ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              |||||||| | |||||  |||      :|||||||||| || |||||| ||| | |   |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| |||||::||:|||||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

RESULT 2
PALY_RHORB
ID   PALY_RHORB     STANDARD;      PRT;   713 AA.
AC   P10248;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Rhodotorula rubra (Yeast) (Rhodotorula mucilaginosa).
OC   Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
OC   Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
OC   Rhodotorula.
OX   NCBI_TaxID=5537;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL Y-15597;
RX   MEDLINE=89083582; PubMed=3205749;
RA   Filpula D., Strausberg R.L., Vaslet C.A., Sykes A., Levy A.;
RT   "Nucleotide sequence of gene for phenylalanine ammonia-lyase from
RT   Rhodotorula rubra.";
RL   Nucleic Acids Res. 16:11381-11381(1988).
RN   [2]
RP   DETAILS.
RC   STRAIN=NRRL Y-15597;
RX   MEDLINE=89083583; PubMed=3205750;
RA   Filpula D., Strausberg R.L., Vaslet C.A., Sykes A., Levy A.;
RT   "cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes
RT   reveal genomic intron deletions.";
RL   Nucleic Acids Res. 16:11382-11382(1988).
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X13094; CAA31486.1; -.
DR   PIR; S01999; S01999.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    217    219       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     218    218       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   713 AA;  76000 MW;  6EB8317CBB037D88 CRC64;

  Query Match             86.8%;  Score 2922.5;  DB 1;  Length 713;
  Best Local Similarity   82.0%;  Pred. No. 4.3e-219;
  Matches  595;  Conservative   33;  Mismatches   85;  Indels   13;  Gaps    5;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||:||||| ||    || |||   |:           | ||| ||||  |||||||  |||
Db          1 MAPSVDSIATSVANSLSNGLHAAAAAN-GGDVHKKTAGAGSLLPTTETTQLDIVERILAD 59

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              |   :||||:||||||||:|| |:| ||  || ||| |||||| || ||| ||  |||||
Db         60 AGATDQIKLDGYTLTLGDVVGAARRGRSVKVADSPHIREKIDASVEFLRTQLDNSVYGVT 119

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
```

```
                ||||||||||||||||||||||||||||||||| | | ||||||||||||||||||||
Db          120 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSMDGFALGRGLENSLPLEVVRGAMTIRV 179

Qy          181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
                ||||||||||:||||||||||||||||||||||||||||||||||||||:||||||||||
Db          180 NSLTRGHSAVRIVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAASITGHPDSKV 239

Qy          241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
                ||  :|  |||  |:||||| ||:|||||||||||||||||||||||||| |||:||||:
Db          240 HV--DG--KIMSAQEAIALKGLQPVVLGPKEGLGLVNGTAVSASMATLALTDAHVLSLLA 295

Qy          301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
                ||||||||||||||||||||||||||||||||||||||||||||||| :||||| |||||
Db          296 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSKYAVHHETEVKVKD 355

Qy          361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
                |||||||||||| |||||||||||||||||||||||||||||||||||:||| ||||| |
Db          356 DEGILRQDRYPLRCSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDLENKMTHHGGAF 415

Qy          421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
                 ||:| ||||||||:||:||::||||||||||||| ||||||||||||||||||||||||
Db          416 MASSVGNTMEKTRLAVALMGKVSFTQLTEMLNAGMNRALPSCLAAEDPSLSYHCKGLDIA 475

Qy          481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
                ||||||||||||||||:|||||||||||:|||||||||||||||||||||||||||||||
Db          476 AAAYTSELGHLANPVSTHVQPAEMGNQAINSLALISARRTAEANDVLSLLLATHLYCVLQ 535

Qy          541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
                |||||||||     |   || ||||  |     |:  ||  |  :|||:| ||||
Db          536 AVDLRAMEFEHTKAFEPMVTELLKQHFGALAT-----AEVEDKVRKSIYKRLQQNNSYDL 590

Qy          601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
                 | ||| || ||| ||| |:      ||||:||||| |||||:|| || || ||  |||
Db          591 EQRWHDTFSVATGAVVEALAG---QEVSLASLNAWKVACAEKAIALTRSVRDSFWAAPSS 647

Qy          661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
                |||| |||||||||||||||:|||||||||:||||||||:|||||||||||| |    |
Db          648 SSPALKYLSPRTRVLYSFVREEVGVKARRGDVYLGKQEVTIGTNVSRIYEAIKSGCIAPV 707

Qy          721 LVKMLA 726
                ||||:|
Db          708 LVKMMA 713
```

```
RESULT 3
PALY_AMAMU
ID   PALY_AMAMU     STANDARD;      PRT;   740 AA.
AC   O93967;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Amanita muscaria (Fly agaric).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Agaricales; Amanitaceae; Amanita.
OX   NCBI_TaxID=41956;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nehls U., Mikolajewski S., Ecke M., Hampp R.;
RT   "Sugar and nitrogen-dependent regulation of an Amanita muscaria
RT   phenylalanine ammonium lyase gene.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ser-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ010143; CAA09013.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    224    226       5-imidazolinone (Ser-Gly)
FT                                (By similarity).
FT   MOD_RES     225    225       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   740 AA;  80156 MW;  B055CFF3D8B7BCEE CRC64;

  Query Match             38.5%;  Score 1294.5;  DB 1;  Length 740;
  Best Local Similarity   42.6%;  Pred. No. 1.1e-92;
  Matches  288;  Conservative  123;  Mismatches  240;  Indels   25;  Gaps    9;

Qy           68 ELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXN--SVYGVTTGFGG 125
                 ::|| :|::  |  |||     |  : :|  :: :: ||   :   ::     ||||::||||
Db           69 KVDGQTLSIAAVAAAARYNAAVELDESPLVKERVRKSQLAIANKVSTGASVYGLSTGFGG 128

Qy          126 SADTRTEDAISLQKALLEHQLCGVLPTSX---DSFXLGRGLENSLPLEVVRGAMTIRVNS 182
                ||||||:  :  | |||:||  |:||||       |     |:    :|||: ||:||
Db          129 SADTRTDKPMLLGFALLQHQHVGILPTSTEPLDVLPLQDANNTSMPEAWIRGAILIRMNS 188

Qy          183 LTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVH- 241
                | |||| :|   ::|  :     |    : |:|||||:||:|||||||||  | |:|  ||:
Db          189 LIRGHSGIRWELIEKMRELLAANVIPVVPLRGSISSGDLSPLSYIAGTIIGNPSIKVYH 248

Qy          242 -VXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
                    | :|  ::: :|| :||  |  || ||::|||  |||:| |||::|  | ||:
Db          249 GPSKSGIRQIGSSKDVLALHNIEPFPLESKEPLGILNGTAFSASVAALALNEAIHLVLLA 308

Qy          301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
                |  ||:  ||::|  |  ||:|  |  ||:| |  ||   |  | |:::|
Db          309 QVCTAMGTEALIGTRASHAPFIHATARPHPGQVECAENIWNLLDGSKLAQLEEHEVRLED 368

Qy          361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQ-STTDNPLIDVENKXTHHGGN 419
                |:  ||||||||||||:|||  : |:|  |   :: |     |||||| |    :||||
Db          369 DKYTLRQDRYPLRTSPQFLGPQIEDIISAFQTVTQECNYLPATDNPLIDGETGESHHGGN 428

Qy          420 FQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDI 479
                ||| || | |||||||  :||| |:| ||::|  ||||| :|| ||||:|| |||||
Db          429 FQAMAVTNAMEKTRLALHHVGKLLFSQSTELVNPAMNRGLPPSVAATDPSLNYHAKGLDI 488

Qy          480 AAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVL 539
                 | ||| :|    | | |||:| ||| ||||||||||||||  |  : :||: |:|::|| :
Db          489 ATAAYVAE----ATPGPTHIQSAEMHNQAVNSLALISARATITSLEVLTSLIASYLYILC 544

Qy          540 QAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYD 599
                ||:||||::  ||          |      ||          :|        |::|  :  |
Db          545 QALDLRALQREFLPGLDIIIREELRSSFGSFLS-SEQMEKLQQNLTSAFEDHLDKTTTMD 603

Qy          600 LEPRWHDAFSXATGTVVEXLSSXXXXXVS-----LAAVNAWKVAXAEKAISLTRXVRXXF 654
                 |     : ::  ::: :          |::|:::: :  |  ::  |    :|  :
Db          604 NTDRMTTMAATSSSVLLQFFTDSGASVPPSSCDLLSSVSSFQSSVATRSSVLMDDLRKEY 663

Qy          655 WXAPSSSSPALXYLSPRTRVLYSFVREELGVKARRGDVF------LGKQEVTIGSNVSRI 708
                   :||  |:  :|| :| |:|  :||:      :  :       ||  ::||||  |:|||
Db          664 IFGDRGPTPASQYIG-KTRPVYQFIRTTIGVRKHGSENYNKFYNGLGVEDVTIGQNISRI 722

Qy          709 YEAIKSGRINXVLVKM 724
                ||:|:  |::   ::| :
```

```
Db        723 YESIRDGKMQSIIVSL 738

RESULT 4
PALY_USTMA
ID   PALY_USTMA     STANDARD;      PRT;   724 AA.
AC   Q96V77;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL1.
OS   Ustilago maydis (Smut fungus).
OC   Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC   Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX   NCBI_TaxID=5270;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim S.-H., Macdonald K., Virmani D., Wake K., Kronstad J.W.,
RA   Ellis B.E.;
RT   "Cloning and disruption of a phenylalanine ammonia-lyase gene from
RT   Ustilago maydis.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC       phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF306551; AAL09388.1; -.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    205    207       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     206    206       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   724 AA;  79277 MW;  02A73C97C34CD1AB CRC64;

  Query Match             35.8%;  Score 1205;  DB 1;  Length 724;
  Best Local Similarity   41.6%;  Pred. No. 9.1e-86;
  Matches  287;  Conservative  114;  Mismatches  245;  Indels   44;  Gaps   15;

Qy         69 LDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQL--XNSVYGVTTGFGGS 126
              :|||:|  :  :| :||  |    |    |  |  | :|| ||: |  :|      |:||: ||||||
Db         41 IDGYNLKIQGLVASARYGHVTRPRPSAETRKRIDDSVQSLIAKLDGGESIYGINTGFGGS 100

Qy        127 ADTRTEDAISLQKALLEHQLCGVLPT---------SXDSFXLG-RGLENSL--PLEVVRG 174
              ||:||  :  :|| |||: | |||||              |    |  |     |:||  |    |||
Db        101 ADSRTANTRALQLALLQMQQCGVLPVPSTFPTGEPSSAPFALPLTDTESSLIMPEAWVRG 160

Qy        175 AMTIRVNSLTRGHSAVRLVLVLEALTN-FLNHGITPIVPLRGTISASGDLSPLSYIAAAIT 233
              |: :|::|| |||| ||  ||: :    || : :||:||:| :||||||||||||:| |:
Db        161 AIVVRLSSLMRGHSGVRWEVLDKMQKLFLQNNVTPVVPVRSSISASGDLSPLSYVAGALA 220

Qy        234 GHPDSKVHVXHEGXEKI-MXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHD 292
              |      |      ::: : | ||  :  : ||   ||| |||:|||| |||:| || ::
Db        221 GQRGIYCFVTDGRGQRVKVTADEACRMHKITPVQYEPKEALGLLNGTAFSASVAGLATYE 280

Qy        293 AHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH- 351
              |  |: |:|  ||: |||: |    || ||:|:: |||| ||: |: || |  | || |  |
Db        281 AENLASLTQLTTAMAVEALKGTDASFAPFIHEIARPHPGQIKSAKFIRAHLSGSRLAEHL 340

Qy        352 -HEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDV- 409
               :|: |    :| | ||||| |:|: ||:|| : |: :|    :  |    ||||||:||
Db        341 ENEKHVLFSEDNGTLRQDRYTLQTASQWVGPGLEDIENAKRSVDFEI-NSTTDNPMIDPY 399

Qy        410 -ENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAA-ED 467
                :   |||||||| |: |  :|| ||||  :||: | |:||::| ||||||: ||:  |
Db        400 DGDGRIHHGGNFQAMAMTNAVEKIRLALCAMGKMTFQQMTELVNPAMNRGLPANLASTPD 459

Qy        468 PSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVL 527
               ||::| ||::|| |: ||||    | |||:|||| ||| ||| ||||||| |:| :| : |
Db        460 LSLNFHAKGINIALASVTSELMFLGNPVSTHVQSAEMANQAFNSLALISGRQTLQAIECL 519

Qy        528 SLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKX 587
              |::  |  ||  |  : ||:|:|:||:::: :|          |    |||          |:  :| |
Db        520 SMIQAWSLYLLCQALDIRALQYKVAEQLPTLILASLHSHFG-EWMDETKQQEIAAQVLKS 578

Qy        588 LXKRLEQTNSYDLEPRWHDAFSXATGTVVEXLSS--XXXXXVSLAAVNAWKVAXAEKAIS 645
              : |||::|:|  ||   |   :  :    |:   :|    |                |  :   |:
Db        579 MSKRLDETSSKDLRDRLVETYQDASSVLVRYFSELPSGGGADPLRNIVKWRATGVADTEK 638

Qy        646 LTRXVRXXFWXAPSSSSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEV------ 699
              : |  |    |    |  :    :  :|  :|:   | |||: |||           ||: :
Db        639 IYRQVTIEFLDNPYACHAS--HLLGKTKRAYEFVRKTLGVPMH------GKENLNEFKGE 690

Qy        700 -----TIGSNVSRIYEAIKSGRINXVLVKM 724
                   |  ||  || :|: |  :   :| ::
Db        691 FEQWNTTGGYVSVIYASIRDGELYNMLSEL 720


RESULT 5
PAL1_ARATH
ID   PAL1_ARATH     STANDARD;      PRT;   725 AA.
AC   P35510; Q9ZQD6;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN   PAL1 OR AT2G37040 OR T1J8.22.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=95195160; PubMed=7888622;
RA   Wanner L.A., Li G., Ware D., Somssich I.E., Davis K.R.;
RT   "The phenylalanine ammonia-lyase gene family in Arabidopsis
RT   thaliana.";
RL   Plant Mol. Biol. 27:327-338(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [3]
```

```
RP   SEQUENCE OF 1-240 FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=93005677; PubMed=2152131;
RA   Ohl S., Hedrick S.A., Chory J., Lamb C.J.;
RT   "Functional properties of a phenylalanine ammonia-lyase promoter from
RT   Arabidopsis.";
RL   Plant Cell 2:837-848(1990).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L33677; AAC18870.1; -.
DR   EMBL; AC006922; AAM15324.1; -.
DR   EMBL; X62747; CAA44609.1; -.
DR   PIR; S52990; S52990.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    211    213       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     212    212       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
FT   CONFLICT    329    329       V -> I (IN REF. 2).
FT   CONFLICT    426    426       R -> A (IN REF. 2).
FT   CONFLICT    612    612       V -> I (IN REF. 2).
SQ   SEQUENCE   725 AA;  78782 MW;  5D65F102D50A2946 CRC64;

  Query Match             28.7%;  Score 965;  DB 1;  Length 725;
  Best Local Similarity   35.5%;  Pred. No. 3.9e-67;
  Matches  258;  Conservative  119;  Mismatches  258;  Indels   92;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRX 99
              |:        : || |:  :|:           | | :||:| |    :   |   |:|   |:   |
Db         39 GAAAEQMKGSHLDEVKRMVAE-FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARA 97

Qy        100 KIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSF 157
               :: | :::    :        ||||||| ::   ||:: ::|| |:           |:   :: ::
Db         98 GVNASSDWVMESMNKGTDSYGVTTGFGATSHRRTKNGVALQKELIRFLNAGIFGSTKET- 156

Qy        158 XLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTIS 217
                       ::||    | || :|:|:| :| | :|   :|||:|:|||: ||| :|||||:
Db        157 ------SHTLPHSATRAAMLVRINTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTIT 210

Qy        218 ASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGL 275
              ||||| |||||| :|| |:||         |  : : | ||   | |:         | ||||| |
Db        211 ASGDLVPLSYIAGLLTGRPNSKA-TGPNG--EALTAEEAFKLAGISSGFFDLQPKEGLAL 267

Qy        276 VNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEV 335
              ||||| : ||::  | : ::||:||:: |:|:   | |  |       |         : || |||
Db        268 VNGTAVGSGMASMVLFETNVLSVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEA 326

Qy        336 ARNIRTLLEGSXF----AVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHA 391
               |  :  :|:|| :          || :    |            :|||| |||||||||  :    :   :|
Db        327 AAVMEHILDGSSYMKLAQKLHEMDPLQKP-----KQDRYALRTSPQWLGPQIE--VIRYA 379

Qy        392 VLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450
               |:|      |  ||||||         |||||  :  :  :|: ||||:  |||| | |  :|:
Db        380 TKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIRAIGKLMFAQFSEL 439

Qy        451 LNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509
               :|   | |||| | |: :|||  |   || :||  |:| |||  :||||||:||| ||  || |
Db        440 VNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDV 499

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569
              ||| |||:|:|:|| |:| |:  |  |  :  ||||| :|    ::
Db        500 NSLGLISSRKTSEAVDILKLMSTTFLVAICQAVDLRHLEENLRQ--------------- 543

Qy        570 XXTXXXXXXELXXKV--------------XKXLXKRL--EQTNSYDLEPRWHDAFSXAT 612
                |      ::  ||                    | | | :   || :| :|           ||
Db        544 --TVKNTVSQVAKKVLTTGVNGELHPSRFCEKDLLKVVDREQVYTYADDP------CSAT 595

Qy        613 GTVVEXLSSXXXXXVSLAAVNA-----------WKVAXAEKAIS--LTRXVRXXFWXAPS 659
                ::: |          | | ||             |:    |: :    | : |         :
Db        596 YPLIQKLRQ---VIVDHALVNGESEKNAVTSIFHKIGAFEEELKAVLPKEVEAARAAYDN 652

Qy        660 SSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRIN 718
               :|    :    |:  || |||||||          ::   |::   :  |      ::: ||  |:|
Db        653 GTSAIPNRIKECRSYPLYRFVREELGT-----ELLTGEKVTSPGEEFDKVFTAICEGKII 707

Qy        719 XVLVKML 725
               ::: |
Db        708 DPMMECL 714


RESULT 6
PAL3_PETCR
ID   PAL3_PETCR     STANDARD;      PRT;   718 AA.
AC   P45729;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 3 (EC 4.3.1.5).
GN   PAL3.
OS   Petroselinum crispum (Parsley) (Petroselinum hortense).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;
OC   Apium clade; Petroselinum.
OX   NCBI_TaxID=4043;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95010141; PubMed=7925471;
RA   Appert C., Logemann E., Hahlbrock K., Schmid J., Amrhein N.;
RT   "Structural and catalytic properties of the four phenylalanine
RT   ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.).";
RL   Eur. J. Biochem. 225:491-499(1994).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X81159; CAA57057.1; -.
```

```
DR   PIR; S48726; S48726.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK     204    206       5-imidazolinone (Ala-Gly)
FT                                 (By similarity).
FT   MOD_RES      205    205       DHA (2,3-DIDEHYDROALANINE)
FT                                 (BY SIMILARITY).
SQ   SEQUENCE   718 AA;  78165 MW;  17786451FFDBDC35 CRC64;

  Query Match             28.5%;  Score 958.5;  DB 1;  Length 718;
  Best Local Similarity   35.2%;  Pred. No. 1.2e-66;
  Matches  264;  Conservative  115;  Mismatches  285;  Indels   87;  Gaps   21;

Qy         14 NGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYS 73
              || || ||             |           : || |: :|:      :|:| :
Db          5 NGTTNG-HANGNGLDLCMKKEDPLNWGVAAEALTGSHLDEVKRMVAE-YRKPVVKLEGET 62

Qy         74 LTLGDVVG-AARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFGGSADTR 130
              ||:   |   :||    |:|  |:|  |  :  | :::   :     ||||||| ::  |
Db         63 LTISQVAAISARDDSGVKVELSEEARAGVKASSDWVMDSMNKGTDSYGVTTGFGATSHRR 122

Qy        131 TEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLE---NSLPLEVVRGAMTIRVNSLTRGH 187
              |:   :|||  |:     |:              |  |  |   |:||   | || :|:|:|:| :|:
Db        123 TKQGGALQKELIRFLNAGI---------FGSGAEAGNNTLPHSATRAAMLVRINTLLQGY 173

Qy        188 SAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEGX 247
              | :|   :|||:| ||||  |||  :|||||:||||| ||||||  :|| |:||  |    |
Db        174 SGIRFEILEAITKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPTGV 232

Qy        248 EKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTA 305
                :    ||  | |:|    | |||||  ||||||  : ||:: | :|::|::|:: ::|
Db        233 --TLSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEVMSA 290

Qy        306 LTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDDEGIL 365
               :  | | |      |    |     : ||  |||  |  :   :|:|| : |    :::    |
Db        291 IFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEMDPLQKP 348

Qy        366 RQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQASAV 425
              :||||  |||||||||| :  ::|  :     :       |  ||||||      ||||||  |  :
Db        349 KQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGSPI 407

Qy        426 XNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAY 484
               :|: ||||:| ||||  |  |  | :|::|    | ||||  |:    :|||  |  || :||  |:|
Db        408 GVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASY 467

Qy        485 TSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQAVDL 544
               |||  ||||||  |||  ||   ||  ||||  |||:|:|:||  ::|  |:   |  |   :  ||:||
Db        468 CSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAIDL 527

Qy        545 RAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKV--------------XKXLX 589
              | :|  |                          ::  :|                  |  |
Db        528 RHLEENLK-----------------STVKNTVSQVAKRVLTMGVNGELHPSRFCEKDLL 569

Qy        590 KRLEQ-----------TNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVA 638
               :  :::          :  :|  |   :    :  :||      :    :|:      |:|
Db        570 RVVDREYIFAYIDDPCSATYPLMQKLRE-------TLVEHALNNGDKERNLSTSIFQKIA 622

Qy        639 XAEKAIS--LTRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSFVREELGVKARRGDVFL 694
                |  :  ·|  :  |       |  |  :||:        |:   ||  ||||||        :
Db        623 AFEDELKALLPKEVETA-RAALESGNPAIPNRIKECRSYPLYKFVREELGT-----EYLT 676

Qy        695 GKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
              |:: :  :|      ::: |:  |   |  |:: |
Db        677 GEKVRSPGEEFEKVFTAMSKGEIIDPLLECL 707

RESULT 7
PALY_PINTA
ID   PALY_PINTA     STANDARD;      PRT;   754 AA.
AC   P52777;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Pinus taeda (Loblolly pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3352;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Needle;
RA   Zhang X.H., Chiang V.L.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U39792; AAA84889.1; -.
DR   PIR; T09777; T09777.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK     206    208       5-imidazolinone (Ala-Gly)
FT                                 (By similarity).
FT   MOD_RES      207    207       DHA (2,3-DIDEHYDROALANINE)
FT                                 (BY SIMILARITY).
SQ   SEQUENCE   754 AA;  82600 MW;  59E97746AA0CA398 CRC64;

  Query Match             28.4%;  Score 957;  DB 1;  Length 754;
  Best Local Similarity   36.9%;  Pred. No. 1.7e-66;
  Matches  252;  Conservative  109;  Mismatches  258;  Indels   64;  Gaps   18;

Qy         69 LDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLX--NSVYGVTTGFGGS 126
               :| |||: ||    ||: :     |:   : ::::|  ::  |:        ||||||| :
Db         64 IEGKSLTISDVAAVARRSQVKVKLDAAAAKSRVEESSNWVLTQMTKGTDTYGVTTGFGAT 123

Qy        127 ADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTRG 186
               :  ||       |||  |:     ||             ||:  || |   :   |  ||:| |:|  :|
Db        124 SHRRTNQGAELQKELIRFLNAGV---------LGKCPENVLSEDTTRAAMLVRTNTLLQG 174

Qy        187 HSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEG 246
               :|  ||  :|| :    ||   :||  :||||||:||||| ||||||   :||  |:|:|     :|
Db        175 YSGVRWDILETVEKLLNAWLTPKLPLRGTITASGDLVPLSYIAGLLTGRPNSRVR-SRDG 233

Qy        247 XEKIMXAREAIALFGLE-PVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTA 305
                |    |    ||:     |||  |   |||||  :||||:|  |::|::  ||::|:|||: ::|
Db        234 IE--MSGAEALKKVGLEKPFELQPKEGLAIVNGTSVGAALASIVCFDANVLALLSEVISA 291

Qy        306 LTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH----HEEEVKVKDD 361
```

```
Db        292 MFCEVMNGKPEFTDPLTHKL-KHHPGQMEAAAIMEYVLDGSSYMKHAAKLHEMNPLQKP- 349

Qy        362 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQ 421

Db        350 ----KQDRYGLRTSPQWLGPQVEIIRSATHMIEREI-NSVNDNPVIDVARDKALHGGNFQ 404

Qy        422 ASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIA 480

Db        405 GTPIGVSMDNLRLSISAIGKLMFAQFSELVNDYYNGGLPSNLSGGPNPSLDYGLKGAEIA 464

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Db        465 MASYTSELLYLANPVTSHVQSAEQHNQDVNSLGLVSARKSAEAIDILKLMLSTYLTALCQ 524

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXK-VXKXLXKRLEQTN--- 596

Db        525 AVDLRHLE----ENMLATVKQIVSQVAKKTLSTGLNGELLPGRFCEKDLLQVVDNEHVFS 580

Qy        597 --------SYDLEPRWHD-----AFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKA 643

Db        581 YIDDPCNASYPLTQKLRNILVEHAFKNAEGEKDPNTSIFNKIPVFEAELK----AQLEPQ 636

Qy        644 ISLTRXVRXXFWXAPSSSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIG 702

Db        637 VSLARE------SYDKGTSPLPDRIQECRSYPLYEFVRNQLGTK-----LLSGTRTISPG 685

Qy        703 SNVSRIYEAIKSGRINXVLVKML 725

Db        686 EVIEVVYDAISEDKVIVPLFKCL 708
```

RESULT 8

```
PALY_MEDSA
ID   PALY_MEDSA     STANDARD;      PRT;   725 AA.
AC   P27990;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS   Medicago sativa (Alfalfa).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
OX   NCBI_TaxID=3879;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Apollo;
RX   MEDLINE=91355933; PubMed=1715786;
RA   Gowri G., Paiva N.L., Dixon R.A.;
RT   "Stress responses in alfalfa (Medicago sativa L.) 12. Sequence
RT   analysis of phenylalanine ammonia-lyase (PAL) cDNA clones and
RT   appearance of PAL transcripts in elicitor-treated cell cultures and
RT   developing plants.";
RL   Plant Mol. Biol. 17:415-429(1991).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X58180; CAA41169.1; -.
DR   PIR; S17444; S17444.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism.
FT   CROSSLNK    211    213       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     212    212       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
FT   DOMAIN       24     27       POLY-ASN.
SQ   SEQUENCE   725 AA;  78865 MW;  19906A2575F64D7D CRC64;

  Query Match             28.4%;  Score 954;  DB 1;  Length 725;
  Best Local Similarity   36.2%;  Pred. No. 2.8e-66;
  Matches  262;  Conservative  106;  Mismatches  252;  Indels  104;  Gaps   21;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFL 108

Db         48 SHLDEVKRMVAE-YRKPVVRLGGETLTISQVAAIAAHDHGVQVDLSESARDGVKASSEWV 106

Qy        109 RXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENS 166

Db        107 MESMNKGTDSYGVTTGFGATSHSRTKQGGALQKELIRFLNAGIFGNGTES-------NHT 159

Qy        167 LPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLS 226

Db        160 LPKTATRAAMLVRINTLLQGYSGIDFEILEAITKPLNKTVTPCLPLRGTITASGDLVPLS 219

Qy        227 YIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGL--EPVVLGPKEGLGLVNGTAVSAS 284

Db        220 YIAGLLTGRPNSK---AHGPSGEVLNAKEAFNLAGINAEFFELQPKEGLALVNGTAVGSG 276

Qy        285 MATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLE 344

Db        277 LASIVLFEANILAVLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILD 335

Qy        345 GSXFAVH----HEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAG-Q 399

Db        336 GSSYVKAAKKLHEIDPLQKP-----KQDRYALRTSPQWLGPLVE--VIRFSTKSIEREIN 388

Qy        400 STTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGL 459

Db        389 SVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLALASIGKLMFAQFSELVNDFYNNGL 448

Qy        460 PSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISAR 518

Db        449 PSNLSASRNPSLDYGFKGAEIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISAR 508

Qy        519 RTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXX 578

Db        509 KTNEAIEILQLMSSTFLIALCQAIDLRHLEENLKN------------------------- 543

Qy        579 ELXXKVXKXLXKRLEQTNSYDLEP------------RWH-----DAFSXATGTVVEXLS 620

Db        544 SVKNTVSQVAKKTLTMGVNGELHPSRFCEKDLLKVVDREHVFAYIDDPCSATYPLSQKLR 603

Qy        621 SXXXXXVSLAAVNA-----------WKVAXAEKAIS--LTRXVRXXFWXAPSSSSPALXY 667

Db        604 Q---VLVDHALVNGESEKNFNTSIFQKIATFEEELKTLLPKEVESA-RTAYESGNPTI-- 657

Qy        668 LSP------RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVL 721

Db        658 --PNKINGCRSYPLYKFVREELGT-----GLLTGENVISPGEECDKLFSAMCQGKIIDPL 710

Qy        722 VKML 725

Db        711 LECL 714
```

```
RESULT 9
PAL1_ORYSA
ID   PAL1_ORYSA     STANDARD;      PRT;   701 AA.
AC   P14717;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RX   MEDLINE=90032676; PubMed=2806257;
RA   Minami E., Ozeki Y., Matsuoka M., Koizuka N., Tanaka Y.;
RT   "Structure and some characterization of the gene for phenylalanine
RT   ammonia-lyase from rice plants.";
RL   Eur. J. Biochem. 185:19-25(1989).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X16099; CAA34226.1; -.
DR   HSSP; P21310; 1B8F.
DR   Gramene; P14717; -.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    189    191       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     190    190       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   701 AA;  75808 MW;  2BF9ADF0BC8E04DE CRC64;

  Query Match             28.3%;  Score 952.5;  DB 1;  Length 701;
  Best Local Similarity   35.9%;  Pred. No. 3.5e-66;
  Matches  254;  Conservative  117;  Mismatches  266;  Indels   71;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDV--VGAARKGRXVRVXDSDEI 97
              |:          : || |:  :|      :     :: | :| :| |   |  |:     |  |   :|
Db         16 GAAAAEMAGSHLDEVKRMVAQ-FREPLVKIQGATLRVGQVAAVAQAKDAARVAVELDEEA 74

Qy         98 RXKIDKSVEFLRXQLXN--SVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXD 155
              | ::  | |::   : :    :|||||||||::  ||:|  :||  || :    |:  |  |
Db         75 RPRVKASSEWILTCIAHGGDIYGVTTGFGGTSHRRTKDGPALQVELLRYLNAGIFGTGSD 134

Qy        156 SFXLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGT 215
                            ::|| | || || :|:|:| :|:| :|   :|||:| || |:|| :||||
Db        135 G--------HTLPSETVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVTPCLPLRGT 186

Qy        216 ISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGL 273
              |:||||| |||||| ||| |:::   :   :| :   : | ||   | |:|        | |||||
Db        187 ITASGDLVPLSYIAGLITGRPNAQA-ISPDGRK--VDAAEAFKLAGIEGGFFTLNPKEGL 243

Qy        274 GLVNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQI 333
              :|||:| :::|    :  ||::|::||: |:|:  | | |              | :  : || |
Db        244 AIVNGTSVGSALAATVMFDANILAVLSEVLSAVFCEVMNGKPEYTDHLTHKL-KHHPGSI 302

Qy        334 EVARNIRTLLEGSXFAVHHEEEVKVKDDEGIL--RQDRYPLRTSPQWLGPLVSDMIHAHA 391
              : |  :  :| || |  | :     || : : :|  :|||| ||||||||| :   :| | |
Db        303 DAAAIMEHILAGSSFMSHAK---KVNEMDPLLKPKQDRYALRTSPQWLGPQI-QVIRA-A 357

Qy        392 VLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450
              |:|      |    |||:|||       |||||| : :  :|:  |||:| |||| |  | :|:
Db        358 TKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSEL 417

Qy        451 LNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509
              :|   | || | || : :||| |  || :|| |:|:||| :||||:| ||| ||  || |
Db        418 VNEFYNNGLTSNLAGSRNPSLDYGFKGTEIAMASYSSELQYLANPITNHVQSAEQHNQDV 477

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569
              ||| |:||:| || |:| |: :|::   : ||||| :|    |                   :   :
Db        478 NSLGLVSARKTLEAVDILKLMTSTYIVALCQAVDLRHLEENIKSSVKNCVTQVAKKVLTM 537

Qy        570 XXTXXXXXXELXXKVXKXLXKRLEQTNSYDLEP--RWHDAFSXATGTVVE-XLSSXXXXX 626
                   |            |           | |   ||   :|      :        :|| |:|
Db        538 NPTGDLSSARFSEKNLLTAIDR-EAVFSYADDPCSANYPLMQKLRAVLVEHALTSGDRRA 596

Qy        627 VSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSSSSP----------------ALXYLSP 670
                 |  :  :|                       ||  ::|                |    :
Db        597 RGLRVLQDHQVRGG----------------APLCAAPGDRGRPRRRRQRTAPVANRIVES 640

Qy        671 RTRVLYSFVREELGVKARRGDVFL-GKQEVTIGSNVSRIYEAIKSGRI 717
              |:  || ||||||         ||| |::   :  |      ::::   |  |::
Db        641 RSFPLYRFVREELGC------VFLTGEKLKSPGEECNKVFLGISQGKL 682

RESULT 10
PALY_TRISU
ID   PALY_TRISU     STANDARD;      PRT;   725 AA.
AC   P45734;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL1.
OS   Trifolium subterraneum (Subterranean clover).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
OX   NCBI_TaxID=3900;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Karridale; TISSUE=Leaf;
RX   MEDLINE=94171061; PubMed=8125321;
RA   Howles P.A., Arioli T., Weinman J.J.;
RT   "Characterization of a phenylalanine ammonia-lyase multigene family
RT   in Trifolium subterraneum.";
RL   Gene 138:87-92(1994).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M91192; AAA17993.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    211    213       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     212    212       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   725 AA;  78998 MW;  70F9925C86622240 CRC64;

  Query Match             28.2%;  Score 949;  DB 1;  Length 725;
  Best Local Similarity   37.2%;  Pred. No. 6.7e-66;
  Matches  252;  Conservative  108;  Mismatches  266;  Indels   52;  Gaps   18;

Qy         69 LDGYSLTLGDVVG-AARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFGG 125
Db         68 LGGETLTISQVAAIAAHDGATVEL--SESARAGVKASSDWVMESMNKGTDSYGVTTGFGA 125

Qy        126 SADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTR 185
Db        126 TSHRRTKQGGALQKELIRFLNAGIFGNGTES-------NHTLPHTATRAAMLVRINTLLQ 178

Qy        186 GHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHE 245
Db        179 GYSGIRFEILEAITKLLNNNITPCLPLRGTITASGDLVPLSYIAGLLTGRSNSK---AHG 235

Qy        246 GXEKIMXAREAIALFGL--EPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQAL 303
Db        236 PSGEMLNAKEAFQLAGINAEFFELQPKEGLALVNGTAVGSGLASIVLFEANILAVLSEVL 295

Qy        304 TALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDDEG 363
Db        296 SAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILHGSAY-VKDAKKLHEMDPLQ 353

Qy        364 ILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAG-QSTTDNPLIDVENKXTHGGNFQA 422
Db        354 KPKQDRYALRTSPQWLGPLIE--VIRFSTKSIEREINSVNDNPLIDVSRNKALHGGNFQG 411

Qy        423 SAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAA 481
Db        412 TPIGVSMDNTRLALASIGKLLFAQFSELVNDFYNNGLPSNLSASRNPSLDYGFKGSEIAM 471

Qy        482 AAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQA 541
Db        472 ASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTKEAIEILQLMSSTFLIALCQA 531

Qy        542 VDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDLE 601
Db        532 IDLRHLEENLKNSVKNTVSQVAKKTLTIGVSGELHPSRFCEKDLLKVVDR-EHVFSYIDD 590

Qy        602 PRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNA-----------WKVAXAEKAIS--LTR 648
Db        591 P------CSATYPLAQKLRQ---VLVDHALVNGESEKNSNTSIFQKIATFEEELKTLLPK 641

Qy        649 XVRXXFWXAPSSSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSR 707
Db        642 EVESARTAYENGNSTIANKINGCRSYPLYKFVREELGT-----SLLTGERVISPGEECDK 696

Qy        708 IYEAIKSGRINXVLVKML 725
Db        697 LFTAMCQGKIIDPLLKCL 714

RESULT 11
PAL1_PRUAV
ID   PAL1_PRUAV     STANDARD;      PRT;   717 AA.
AC   O64963;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN   PAL1.
OS   Prunus avium (Cherry).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX   NCBI_TaxID=42229;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Summit;
RA   Wiersma P.A., Wu Z.;
RT   "A full-length cDNA for phenylalanine ammonia-lyase cloned from ripe
RT   Sweet Cherry fruit (Prunus avium).";
RL   (In) Plant Gene Register PGR98-184.
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF036948; AAC78457.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    203    205       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     204    204       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   717 AA;  77999 MW;  B84DF90FA0BF60B3 CRC64;

  Query Match             28.1%;  Score 947;  DB 1;  Length 717;
  Best Local Similarity   35.8%;  Pred. No. 9.5e-66;
  Matches  250;  Conservative  112;  Mismatches  285;  Indels   52;  Gaps   17;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRX 99
Db         31 GVAAETLKGSHLDEVKRMVAE-YRKPVVKLGGESLTISQVAAIATHDSGVKVELSESARA 89

Qy        100 KIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSF 157
Db         90 GVKASSDWVMDSMSKGTDSYGVTTGFGATSHRRTKQGAALQKELIRFLNAGVFGSTKES- 148

Qy        158 XLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTIS 217
Db        149 ------GHTLPHQATRAAMLVRINTLLQGYSGIRFEILEVITKFLNNNVTPCLPLRGTIT 202
```

```
Qy        218 ASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGL 275
              ||||| ||||||  :|| |:||  |  :|   : : | ||     |:      | ||||| |
Db        203 ASGDLVPLSYIAGMLTGRPNSKA-VGPDG--QTLSAAEAFEFVGINSGFFELQPKEGLAL 259

Qy        276 VNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEV 335
              ||||||  : :|:  | |  ::|:|:|||: |:|:  | | |     |     |     : || |||
Db        260 VNGTAVGSGLASTVLFDTNILALLSEILSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEA 318

Qy        336 ARNIRTLLEGSXFAVH----HEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHA 391
              |  :  :|:|| :      ||::   |        :||||  ||||||||||  :    :  ::
Db        319 AAIMEHILDGSSYVKAAKKLHEQDPLQKP-----KQDRYALRTSPQWLGPQIE--VIRYS 371

Qy        392 VLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450
              |:|   |  |||||||         |||||| : :   :|: ||||:| ||||  |  | :|:
Db        372 TKSIEREIDSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLAIASIGKLMFAQFSEL 431

Qy        451 LNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509
              :|   | ||||  |:    :|||  |   || :|| |:| ||| |||||| ||| ||  || |
Db        432 VNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDV 491

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569
              ||| |||:|:||||  |:| |:  :| |   : ||:||| :|     :          :  :
Db        492 NSLGLISSRKTAEAVDILKLMSSTFLVALCQAIDLRHLEENLRNTVKNTVSQVAKR---T 548

Qy        570 XXTXXXXXXELXXKVXKXLXKRLEQTNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXXVSL 629
                    |        | | |  :::   :   :  |   ||  ::: |
Db        549 LTTGVNGELHPSRFCEKDLLKVVDREYVF----AYIDDPCSATYPLMQKLRQVLVEHALT 604

Qy        630 AAVNAWKVAXA--EKAISLTRXVRXXFWXAPSSSSPALXYLSP---------RTRVLYSF 678
                 |   : :  :|  ::     ::           |:  ||  |          |:  ||  |
Db        605 NGENEKNASTSIFQKIVAFEEELKVLLPKEVDSARAALDSGSAGVPNRITECRSYPLYKF 664

Qy        679 VREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRI 717
              |||||     :    |::  : |     :::  ||  |:|
Db        665 VREELG-----AEYLTGEKVRSPGEECDKVFTAICEGKI 698
```

```
RESULT 12
PAL1_PETCR
ID   PAL1_PETCR     STANDARD;      PRT;   716 AA.
AC   P24481;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN   PAL1.
OS   Petroselinum crispum (Parsley) (Petroselinum hortense).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;
OC   Apium clade; Petroselinum.
OX   NCBI_TaxID=4043;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89356635; PubMed=2767049;
RA   Lois R., Dietrich A., Hahlbrock K., Schulz W.;
RT   "A phenylalanine ammonia-lyase gene from parsley: structure,
RT   regulation and identification of elicitor and light responsive
RT   cis-acting elements.";
RL   EMBO J. 8:1641-1648(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kang X., Logemann E., Hahlbrock K.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   ACTIVE SITE, AND MUTAGENESIS OF SER-203 AND SER-210.
RX   MEDLINE=94326930; PubMed=8050576;
RA   Schuster B., Retey J.;
RT   "Serine-202 is the putative precursor of the active site
RT   dehydroalanine of phenylalanine ammonia lyase. Site-directed
RT   mutagenesis studies on the enzyme from parsley (Petroselinum crispum
RT   L.).";
RL   FEBS Lett. 349:252-254(1994).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X15473; CAA33500.1; -.
DR   EMBL; X16772; CAA34715.1; -.
DR   EMBL; Y07654; CAA68938.1; -.
DR   PIR; S04463; S04463.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    202    204       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     203    203       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
FT   MUTAGEN     203    203       S->A: COMPLETE LOSS OF ACTIVITY.
FT   MUTAGEN     210    210       S->A: NO LOSS OF ACTIVITY.
SQ   SEQUENCE   716 AA;  77828 MW;  23D82FA1AC6FBB4A CRC64;

  Query Match             28.1%;  Score 946.5;  DB 1;  Length 716;
  Best Local Similarity   36.7%;  Pred. No. 1e-65;
  Matches  254;  Conservative  110;  Mismatches  286;  Indels   43;  Gaps   18;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVG-AARKGRXVRVXDSDEIRXKIDKSVEF 107
               : || |:  :|:           :| | :||:  |    :||  |  |  |  |  |: |   :  | ::
Db         40 SHLDEVKKMVAE-YRKPVVKLGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDW 98

Qy        108 LRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLEN 165
               :     :        |||||||| ::  ||:  :|| |:      |:             | | :|
Db         99 VMDSMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI---------FGNGSDN 149

Qy        166 SLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL 225
               :||     | || :|:|:| :|:| :|  :|||:| |||  ||| :||||||:||||| ||
Db        150 TLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPL 209

Qy        226 SYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSA 283
              ||||  :|| |:||  |   |     |:     ||  | |:        | |||||  ||||||  :
Db        210 SYIAGLLTGRPNSKA-VGPTGV--ILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGS 266

Qy        284 SMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLL 343
              ||:: | :|::|::|:: ::|:  | | |      |     |      : || ||| |  :  :|
Db        267 GMASMVLFEANILAVLAEVMSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHIL 325

Qy        344 EGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTD 403
              :|| :  |     :::   |       :||||  ||||||||||  : ::|  :    :       |   |
Db        326 DGSAY-VKAAQKLHEMDPLQKPKQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINSVND 383

Qy        404 NPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL 463
              ||||||         ||||||  : :   :|:  ||||:| ||||  |  | :|::|   | ||||  |
Db        384 NPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNL 443

Qy        464 A-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAE 522
                :  :|||  |   || :|| |:| |||  |||||| ||| ||  ||  |||| |||:|:|:|
```

```
Db        444 SGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSE 503

Qy        523 ANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXX 582

Db        504 AVEILKLMSTTFLVGLCQAIDLRHLEENLKSTVKNTVSSVAKRVLTMGVNGELHPSRFCE 563

Qy        583 KVXKXLXKRLEQTNSYDLEP--RWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXA 640

Db        564 KDLLRVVDR-EYIFAYIDDPCSATYPLMQKLRQTLVEHALKNGDNERNLSTSIFQKIATF 622

Qy        641 EKAIS--LTRXVRXXFWXAPSSSSPALXYLSP------RTRVLYSFVREELGVKARRGDV 692

Db        623 EDELKALLPKEVESA-RAALESGNPAI----PNRIEECRSYPLYKFVRKELGT-----EY 672

Qy        693 FLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725

Db        673 LTGEKVTSPGEEFEKVFIAMSKGEIIDPLLECL 705


RESULT 13
PAL2_PETCR
ID   PAL2_PETCR     STANDARD;      PRT;   716 AA.
AC   P45728;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
GN   PAL2.
OS   Petroselinum crispum (Parsley) (Petroselinum hortense).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;
OC   Apium clade; Petroselinum.
OX   NCBI_TaxID=4043;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95010141; PubMed=7925471;
RA   Appert C., Logemann E., Hahlbrock K., Schmid J., Amrhein N.;
RT   "Structural and catalytic properties of the four phenylalanine
RT   ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.).";
RL   Eur. J. Biochem. 225:491-499(1994).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; X81158; CAA57056.1; -.
DR   PIR; S48725; S48725.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    202    204       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     203    203       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   716 AA;  77964 MW;  F74D7BD881BEEE66 CRC64;

  Query Match             28.1%;  Score 946.5;  DB 1;  Length 716;
  Best Local Similarity   35.5%;  Pred. No. 1e-65;
  Matches  254;  Conservative  110;  Mismatches  264;  Indels   87;  Gaps   19;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVG-AARKGRXVRVXDSDEIRXKIDKSVEF 107

Db         40 SHLDEVKKMVAE-YRKPVVKLGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDW 98

Qy        108 LRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLEN 165

Db         99 VMDSMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI---------FGNGSDN 149

Qy        166 SLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL 225

Db        150 TLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPL 209

Qy        226 SYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSA 283

Db        210 SYIAGLLTGRPNSKA-VGPTGV--ILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGS 266

Qy        284 SMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLL 343

Db        267 GMASMVLFEANILAVLAEVMSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHIL 325

Qy        344 EGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTD 403

Db        326 DGSAY-VKAAQKLHEMDPLQKPKQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINSVND 383

Qy        404 NPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL 463

Db        384 NPLIDVSRNKAIHGGNFQGTPIGMSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNL 443

Qy        464 A-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAE 522

Db        444 SGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSE 503

Qy        523 ANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHF-----GXXXTXXXXX 577

Db        504 AVEILKLMSTTFLVGLCQAIDLRHLEENLKSTVKNTVSSVAKRVLTMGVNGELHPSRFCE 563

Qy        578 XELXXKKVXKX---------------LXKRLEQT------NSYDLEPRWHDAFSXATGTVV 616

Db        564 KDLLRFVDREYIFAYIDDPCSATYPLMQKLRQTLVEHALKNGDNERNMNTSIFQKIATFE 623

Qy        617 EXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSSSSPALXYLSP------ 670

Db        624 DELK-----------------ALLPKEVESAR-------AALESGNPAI----PNRIEEC 655

Qy        671 RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725

Db        656 RSYPLYKFVRKELGI-----EYLTGEKVTSPGEEFDKVFIAMSKGEIIDPLLECL 705


RESULT 14
PAL1_SOYBN
ID   PAL1_SOYBN     STANDARD;      PRT;   713 AA.
AC   P27991;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN   PAL1.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. T225; TISSUE=Leaf;
```

```
RX   MEDLINE=92190550; PubMed=1799682;
RA   Frank R.L., Vodkin L.O.;
RT   "Sequence and structure of a phenylalanine ammonia-lyase gene from
RT   Glycine max.";
RL   DNA Seq. 1:335-346(1991).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X52953; CAA37129.1; -.
DR   PIR; S22991; S22991.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    199    201       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     200    200       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   713 AA;  77744 MW;  9D71EF1CC230216A CRC64;

  Query Match             28.1%;  Score 946;  DB 1;  Length 713;
  Best Local Similarity   36.6%;  Pred. No. 1.1e-65;
  Matches  271;  Conservative  105;  Mismatches  293;  Indels   72;  Gaps   23;

Qy         14 NGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYS 73
              ||  ||        |             | |            : || |:  :|:       | | :
Db          5 NGHQNGSFCLSTAKGNNDPLNWGAAA----EAMKGSHLDEVKRMVAE-YRKPVVRLGGET 59

Qy         74 LTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRT 131
              ||:  |    |       | |  |:  |   :  | |::   :  |   ||||||| ::  ||
Db         60 LTIAQVAAVAGHDHGVAVELSESAREGVKASSEWVMNSMNNGTDSYGVTTGFGATSHRRT 119

Qy        132 EDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENS--LPLEVVRGAMTIRVNSLTRGHSA 189
               :  :|||  |:        |:             |  |  |:|  ||   | ||  :|:|:|:|  :|:|
Db        120 KQGGALQKELIRFLNAGI---------FGNGTESSHTLPHTATRAAMLVRINTLLQGYSG 170

Qy        190 VRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEK 249
              :|  :|||:|  ||: :|| :  |||||:|||||  ||||||  :||  |:||  |       |   :
Db        171 IRFEILEAITKLLNNNVTPCLDLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPSG--E 227

Qy        250 IMXAREAIALFGL--EPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTALT 307
              :: |:||  |  :  |   | ||||| ||||||  : :|::  |  :|::|::||: |:|:
Db        228 VLNAKEAFELASINSEFFELQPKEGLALVNGTAVGSGLASMVLFEANILAVLSEVLSAIF 287

Qy        308 VEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH----HEEEVKVKDDEG 363
               |  |  |      |     |     :  |||  |  :  :|:||   |         ||  :    |
Db        288 AEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSYMKAAKKLHEIDPLQKP--- 343

Qy        364 ILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAG-QSTTDNPLIDVENKXTHHGGNFQA 422
                :||||  ||||||||||:       :  |  :  |:|        |  |||||||           |||||
Db        344 --KQDRYALRTSPQWLGPLIE--VIRFSTKSIEREINSVNDNPLIDVSRNKALHGGNFQG 399

Qy        423 SAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAA 481
               :  :  :|:  ||||| ||||  |  |  :|::|    |  ||||  |  |:  :|||  |    ||  :||
Db        400 TPIGVSMDNTRLALASIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKGAEIAM 459

Qy        482 AAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQA 541
              |:|  |||  :|||||||||  ||    ||  ||||  |||:|:|  ||  ::|  |:  :|  |    :  ||
Db        460 ASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTNEAIEILKLMSSTFLIALCQA 519

Qy        542 VDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRL--EQTNSYD 599
              :|||  :|     |                :     :                 |  |  |  :    |   ||
Db        520 IDLRHLEENLKNSVKNTVSQVSKR---ILTTGVNGELHPSRFCEKDLLKVVDREYIFSYI 576

Qy        600 LEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNA-----------WKVAXAEKAIS--L 646
               :|         ||  :::  |            |    |  |||                   |:|   |:  :    |
Db        577 DDP------CSATYPLMQKLRQ---VLVDHALVNAECEKDVNSSIFQKIAIFEEELKNLL 627

Qy        647 TRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSN 704
                :  |.       |    |    |:               |:   ||  |||||||          :    |::   :  |
Db        628 PKEVEGA-RAAYESGKAAIPNKIQECRSYPLYKFVREELGT-----GLLTGEKVRSPGEE 681

Qy        705 VSRIYEAIKSGRINXVLVKML 725
                .::: |:   |:|     |:: |
Db        682 FDKLFTAMCQGKIIDPLMECL 702

RESULT 15
PAL4_ARATH
ID   PAL4_ARATH     STANDARD;      PRT;   707 AA.
AC   Q9SS45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   AT3G10340 OR F14P13.6.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
CC   -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC       first reaction in the biosynthesis from L-phenylalanine of a wide
```

```
CC       variety of natural products based on the phenylpropane skeleton.
CC   -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC   -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC       which is formed autocatalytically by cyclization and dehydration
CC       of residues Ala-Ser-Gly (By similarity).
CC   -!- SIMILARITY: Belongs to the PAL / histidase family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AC009400; AAF02809.1; -.
DR   HSSP; P21310; 1B8F.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase; Phenylpropanoid metabolism; Multigene family.
FT   CROSSLNK    194    196       5-imidazolinone (Ala-Gly)
FT                                (By similarity).
FT   MOD_RES     195    195       DHA (2,3-DIDEHYDROALANINE)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   707 AA;  76919 MW;  B67743C941B50B9B CRC64;

  Query Match             28.0%;  Score 943.5;  DB 1;  Length 707;
  Best Local Similarity   37.0%;  Pred. No. 1.7e-65;
  Matches  252;  Conservative   98;  Mismatches  282;  Indels   49;  Gaps   16;

Qy         65 DXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTG 122

Db         45 EAVKLGGETLTIGQVAAVARGGGGSTVELAEEARAGVKASSEWVMESMNRGTDSYGVTTG 104

Qy        123 FGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRG-LENSLPLEVVRGAMTIRVN 181

Db        105 FGATSHRRTKQGGALQNELIRFLNAGI-------FGPGAGDTSHTLPKPTTRAAMLVRVN 157

Qy        182 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVH 241

Db        158 TLLQGYSGIRFEILEAITKLLNHEITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA- 216

Qy        242 VXHEGXEKIMXAREAIALFGLEPVV-LGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Db        217 VGPSG--ETLTASEAFKLAGVSSFFELQPKEGLALVNGTAVGSGLASTVLFDANILAVLS 274

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFA----VHHEEEV 356

Db        275 EVMSAMFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSYVKEAQLLHEMDP 333

Qy        357 KVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHH 416

Db        334 LQKP-----KQDRYALRTSPQWLGPQI-EVIRAATKMIEREINSVNDNPLIDVSRNKALH 387

Qy        417 GGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCK 475

Db        388 GGNFQGTPIGVAMDNSRLAIASIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFK 447

Qy        476 GLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHL 535

Db        448 GAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSTTYL 507

Qy        536 YCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQT 595

Db        508 VALCQAVDLRHLEENLKKAVKSAVSQVAKRVLTVGANGELHPSRFTERDVLQVVDR-EYV 566

Qy        596 NSYDLEP----------RWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAIS 645

Db        567 FSYADDPCSLTYPLMQKLRHILVDHALADPEREANSATSVFHKIGAFE------AELKLL 620

Qy        646 LTRXV-RXXFWXAPSSSSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSN 704

Db        621 LPKEVERVRVEYEEGTSAIANRIKECRSYPLYRFVRDELNT-----ELLTGENVRSPGEE 675

Qy        705 VSRIYEAIKSGRINXVLVKML 725

Db        676 FDKVFLAISDGKLIDPLLECL 696

Search completed: September  9, 2004, 10:04:42
Job time : 19.5685 secs
```

```
OM protein - protein search, using sw model

Run on:         September  9, 2004, 09:57:28 ; Search time 88.3651 Seconds
                                          (without alignments)
                                          2592.269 Million cell updates/sec

Title:          US-09-939-408A-21
Perfect score:  3365
Sequence:       1 MAPSLDSIATSXANGXXNGX..........RIYEAIKSGRINXVLVKMLA 726

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_25:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	969	28.8	725	10	Q94AN1	Q94an1 arabidopsis
2	967	28.7	711	10	Q8W2E4	Q8w2e4 lactuca sat
3	965.5	28.7	707	10	Q7X7G3	Q7x7g3 oryza sativ
4	963	28.6	671	10	Q84VE0	Q84ve0 oryza sativ
5	958.5	28.5	703	10	Q8VXG7	Q8vxg7 zea mays (m
6	955	28.4	711	10	Q9AXI5	Q9axi5 pharbitis n
7	955	28.4	730	10	Q9M567	Q9m567 rubus idaeu
8	954	28.4	708	10	Q94EN0	Q94en0 rehmannia g
9	944.5	28.1	681	10	Q8RUZ3	Q8ruz3 pinus sylve
10	944.5	28.1	715	10	Q852S1	Q852s1 daucus caro
11	943.5	28.0	681	10	Q8SAS6	Q8sas6 pinus sylve
12	942.5	28.0	716	10	Q9MAX1	Q9max1 catharanthu
13	941.5	28.0	681	10	Q8RV49	Q8rv49 pinus sylve
14	940	27.9	717	10	Q8RWP4	Q8rwp4 arabidopsis
15	939	27.9	713	10	Q8GZR8	Q8gzr8 lactuca sat
16	937.5	27.9	681	10	Q8SAS7	Q8sas7 pinus sylve
17	931	27.7	710	10	Q94C45	Q94c45 manihot esc
18	929.5	27.6	716	10	Q9ATN7	Q9atn7 agastache r
19	926.5	27.5	721	10	Q9XFX5	Q9xfx5 citrus clem
20	922	27.4	687	10	Q94F90	Q94f90 manihot esc
21	919.5	27.3	727	10	Q7XJC3	Q7xjc3 pinus pinas
22	917.5	27.3	727	10	Q7XJC4	Q7xjc4 pinus pinas
23	915.5	27.2	714	10	Q7X6M8	Q7x6m8 oryza sativ
24	914	27.2	711	10	Q8H6V5	Q8h6v5 populus tre
25	911.5	27.1	714	10	Q8H6V6	Q8h6v6 populus tre
26	910	27.0	717	10	Q8H6W0	Q8h6w0 coffea cane
27	904.5	26.9	712	10	Q94F89	Q94f89 manihot esc
28	900	26.7	655	10	Q9XGR3	Q9xgr3 vigna ungui
29	898.5	26.7	549	10	O04876	O04876 hordeum vul
30	895.5	26.6	718	10	Q9XFX6	Q9xfx6 citrus clem
31	893	26.5	703	10	Q84LI2	Q84li2 phalaenopsi
32	893	26.5	715	10	O24266	O24266 populus kit
33	892.5	26.5	619	10	Q39545	Q39545 cucumis mel
34	887	26.4	619	10	Q8H6V9	Q8h6v9 coffea cane
35	868.5	25.8	618	10	Q9FS00	Q9fs00 dianthus ca
36	763	22.7	396	10	Q8VWP6	Q8vwp6 beta vulgar
37	736.5	21.9	544	10	Q9FT38	Q9ft38 triticum ae
38	735	21.8	532	10	O81647	O81647 capsicum ch
39	717	21.3	380	10	Q84UZ7	Q84uz7 ginkgo bilo
40	715	21.2	362	10	Q9FUI7	Q9fui7 rubus idaeu
41	703	20.9	363	10	Q9FUI8	Q9fui8 rubus idaeu
42	612.5	18.2	497	10	O04869	O04869 hordeum vul
43	604.5	18.0	475	10	O04875	O04875 hordeum vul
44	600.5	17.8	436	10	Q9LEP1	Q9lep1 betula verr
45	596.5	17.7	435	10	Q9M3U1	Q9m3u1 betula verr

ALIGNMENTS

```
RESULT 1
Q94AN1
ID   Q94AN1     PRELIMINARY;     PRT;   725 AA.
AC   Q94AN1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia lyase (Hypothetical protein) (EC 4.3.1.5).
GN   AT2G37040 OR PAL1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA   Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA   Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA   Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA   Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA   Ecker J.R., Theologis A.;
RT   "Full Length cDNA of gene T1J8.22/At2g37040 (GI:6598547).";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,
RA   Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
RA   Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,
RA   Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
RA   Satou M., Seki M., Shinn P., Tang C.C., Toroumi M., Wong C., Wu H.C.,
RA   Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,
RA   Davis R.W.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
```

```
RA   Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
RA   Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
RA   Southwick A., Davis R.W., Ecker J.R., Theologis A.;
RT   "Arabidopsis Open Reading Frame (ORF) Clones.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Cochrane F.C., Kim M.R., Cardenas C.L., Davin L.B., Lewis N.G.;
RT   "Putative PAL multigene family in Arabidopsis thaliana: Kinetic
RT   Characterization of Recombinant Proteins and Knock-out Mutant
RT   Characterization.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY045919; AAK76593.1; -.
DR   EMBL; BT003330; AAO29949.1; -.
DR   EMBL; AY079363; AAL85094.1; -.
DR   EMBL; AY303128; AAP59438.1; -.
DR   PIR; G84787; G84787.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Hypothetical protein; Lyase.
SQ   SEQUENCE   725 AA;  78725 MW;  02626B3B2DEFE9CE CRC64;

  Query Match            28.8%;  Score 969;  DB 10;  Length 725;
  Best Local Similarity  35.5%;  Pred. No. 8.8e-68;
  Matches  258;  Conservative  120;  Mismatches  257;  Indels   92;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRX 99

Db         39 GAAAEQMKGSHLDEVKRMVAE-FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARA 97

Qy        100 KIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSF 157

Db         98 GVNASSDWVMESMNKGTDSYGVTTGFGATSHRRTKNGVALQKELIRFLNAGIFGSTKET- 156

Qy        158 XLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTIS 217

Db        157 ------SHTLPHSATRAAMLVRINTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTIT 210

Qy        218 ASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGL 275

Db        211 ASGDLVPLSYIAGLLTGRPNSKA-TGPNG--EALTAEEAFKLAGISSGFFDLQPKEGLAL 267

Qy        276 VNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEV 335

Db        268 VNGTAVGSGMASMVLFETNVLSVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEA 326

Qy        336 ARNIRTLLEGSXF----AVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHA 391

Db        327 AAIMEHILDGSSYMKLAQKLHEMDPLQKP-----KQDRYALRTSPQWLGPQIE--VIRYA 379

Qy        392 VLSLEAG-QSTTDNPLIDVENXXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450

Db        380 TKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL 439

Qy        451 LNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509

Db        440 VNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDV 499

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569

Db        500 NSLGLISSRKTSEAVDILKLMSTTFLVAICQAVDLRHLEENLRQ---------------- 543

Qy        570 XXTXXXXXXXELXXKV---------------XKXLXKRL--EQTNSYDLEPRWHDAFSXAT 612

Db        544 --TVKNTVSQVAKKVLTTGVNGELHPSRFCEKDLLKVVDREQVYTYADDP------CSAT 595

Qy        613 GTVVEXLSSXXXXXVSLAAVNA----------WKVAXAEKAIS--LTRXVRXXFWXAPS 659

Db        596 YPLIQKLRQ---VIVDHALINGESEKNAVTSIFHKIGAFEEELKAVLPKEVEAARAAYDN 652

Qy        660 SSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRIN 718

Db        653 GTSAIPNRIKECRSYPLYRFVREELGT-----ELLTGEKVTSPGEEFDKVFTAICEGKII 707

Qy        719 XVLVKML 725

Db        708 DPMMECL 714


RESULT 2
Q8W2E4
ID   Q8W2E4      PRELIMINARY;      PRT;   711 AA.
AC   Q8W2E4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
OS   Lactuca sativa (Garden lettuce).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;
OC   Lactuca.
OX   NCBI_TaxID=4236;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Campos-Vargas R., Nonogaki H., Suslow T., Saltveit M.;
RT   "Characterization of phenylalanine ammonia-lyase (PAL) gene in wounded
RT   lettuce leaf tissue.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF299330; AAL55242.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   711 AA;  77363 MW;  3120AB2877C29513 CRC64;

  Query Match            28.7%;  Score 967;  DB 10;  Length 711;
  Best Local Similarity  36.9%;  Pred. No. 1.2e-67;
  Matches  263;  Conservative  108;  Mismatches  262;  Indels   80;  Gaps   20;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVG--AARKGRXVRVXDSDEIRXKIDKSVE 106

Db         32 SHLDEVKKMVAE-FRKPVVKLGGETLTVSQVAGIAAANDSDTVKVELSEAARAGVKASSD 90

Qy        107 FLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLE 164

Db         91 WVMESMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI---------FGNGTE 141

Qy        165 NS--LPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDL 222

Db        142 TSHTLPHSATRAAMIVRINTLLQGYSGIRFEILEAITKFLNNNITPCLPLRGTITASGDL 201

Qy        223 SPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTA 280

Db        202 VPLSYIAGLLTGRPNSKA-VGPTG--EVLNAEKAFAAAGVEGGFFELQPKEGLALVNGTA 258

Qy        281 VSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIR 340
```

```
                 | : ||:: | ||::|:|||: |:|:  | | |     |     |      : || ||| |   :
Db       259 VGSGMASMVLFDANVLALLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIME 317

Qy       341 TLLEGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQS 400
              :|:|| : |   ::|     |       :|||| ||||||||| : ::| :    :       |
Db       318 YILDGSDY-VKAAQKVHEMDPLQKPKQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINS 375

Qy       401 TTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLP 460
                |||||        |||||  : :   :|: ||||:| ||||  |  | :|::|    | |||
Db       376 VNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLP 435

Qy       461 SCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARR 519
              | |:  :|||  |  || :|| |:| |||  ||||| |||  ||   || |||| |||||:
Db       436 SNLSGGRNPSLDYGFKGGEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISARK 495

Qy       520 TAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHF-----GXXXTXX 574
              |||  |:| |: :|:|   |::|||  :|            :     :          |
Db       496 TAEAVDILKLMSSTYLVALCQSIDLRHLEENMKSTVKNTVSQVAKKVLTMGVNGELHPSR 555

Qy       575 XXXXELXXXVXKX--------------LXKRLEQT------NSYDLEPRWHDAFSXATG 613
                   :|   | :               | ::| |      |: :  |    : :
Db       556 FCEKDLLRVVDREYVFAYIDDVCSGTYPLMQKLRQVLVDHALNNGETEKNTNTSIFQKIA 615

Qy       614 TVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSSSSPALXYLSPRTR 673
              |  | |        :    |    ::|      :|:   ::                 : |:
Db       616 TFEEELK-----VLLPKEVEGVRIAYENDTLSIPNRIK----------------ACRSY 653

Qy       674 VLYSFVREELGVKARRGDVFLGKQEVTI-GSNVSRIYEAIKSGRINXVLVKML 725
              || |||||||     ||    ||  ::||    |  |::  |:  |:|    |:: |
Db       654 PLYRFVREELG----RG--FLTGEKVTSPGEEFDRVFTAMCKGQIIDPLLECL 700
```

RESULT 3
Q7X7G3
ID   Q7X7G3   PRELIMINARY;   PRT;   707 AA.
AC   Q7X7G3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   OSJNBa0073E02.12 protein (OSJNBb0061C13.1 protein).
GN   OSJNBA0073E02.12 OR OSJNBB0061C13.1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA   Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA   Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA   Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA   Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA   Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA   Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA   Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA   Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL731616; CAE05452.1; -.
DR   EMBL; AL731629; CAE05619.1; -.
SQ   SEQUENCE   707 AA; 76038 MW; 841E4C46D077B1AA CRC64;

```
  Query Match             28.7%;  Score 965.5;  DB 10;  Length 707;
  Best Local Similarity   37.1%;  Pred. No. 1.6e-67;
  Matches  257;  Conservative  122;  Mismatches  273;  Indels   41;  Gaps   20;

Qy        49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDV--VGAARKGRXVRVXDSDEIRXKIDKSVE 106
              : || |:  :|     :    :::| || :| |  | ||:     |  |   :| | :: | |
Db        29 SHLDEVKRMVAQ-SREAVVKIEGSSLRVGQVAAVSAAKDASGVVVELDEEARPRVKASSE 87

Qy       107 FLRXQLXN--SVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLE 164
              ::   : :   :|||||||||::  ||:|   :||  || |   |:     |
Db        88 WILNCIAHGGDIYGVTTGFGGTSHRRTKDGQALQVELLRHLNAGIFGNGSDG-------- 139

Qy       165 NSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSP 224
              |||| || | || :|:|:| :|:| :| :|   :|||:   :|  |::| :|||||:||||| |
Db       140 NSLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLINTGVSPCLPLRGTITASGDLVP 199

Qy       225 LSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVS 282
              |||||  |||  |:::   |   :|   |  |  ||    :  |::      ||||| :||||:|
Db       200 LSYIAGLITGRPNAQA-VTVDG--KKVDAAEAFKIAGIQGGFFRLEPKEGLAIVNGTSVG 256

Qy       283 ASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTL 342
              :::| : |:||::|::||: |:|:  | | |                | : : ||  || |   :   :
Db       257 SALAAMVLYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKL-KHHPGSIEAAAIMEHI 315

Qy       343 LEGSXFAVHHEEEVKVKDDEGIL--RQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAG-Q 399
              | || |  | :     || : : :|   :|||| ||||||||| : ::| |  |  |:|
Db       316 LAGSAFMPHAQ---KVNEVDPLLKPKQDRYALRTSPQWLGPQI-EVIRA-ATKSIEREVN 370

Qy       400 STTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGL 459
              |  |||:|||          |||||  : :   :|: ||||:| ||||  |  | :|::|    | ||
Db       371 SVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNTRLAIANIGKLMFAQFSELVNEFYNNGL 430

Qy       460 PSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISAR 518
              | ||  : :||| |  || :|| |:| |||  ||||| |||  ||   || |||| |||:||
Db       431 TSNLAGSRNPSLDYGFKGTEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLVSAR 490

Qy       519 RTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXX 578
              :|||| |:| |: :|:|   : |||||| :|     |               :   :
Db       491 KTAEAVDILKLMSSTYLVALCQAVDLRHLEENLKSAVKNCVTTVAKKVLTTGPAGGLHSA 550

Qy       579 ELXXKVXKXLXKRLEQTNSYDLEP--RWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWK 636
                   |        |  |    || :|        :      :   :||    :            :    |
Db       551 RFSEKALLTAIDR-EAVYSYADDPCSANYPLMTKIRAVLVEHALANGPAEKDDGSSVFSK 609

Qy       637 VAXAEKAI--SLTRXVRXXFWXAPSSSSPALXYL-SPRTRVLYSFVREELGVKARRGDVF 693
              :   |: :  :| | :       : ::|   :     |:  || |||||||          |:
Db       610 ITAFEEELREALPREMEAARVAFETGTAPITNRIKESRSFPLYRFVREELGC------VY 663

Qy       694 L-GKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
              | |:: :  |   :::: ||    ::    ::: |
Db       664 LTGEKLKSPGEECNKVFLAISERKLIDPMLECL 696
```

RESULT 4
Q84VE0
ID   Q84VE0   PRELIMINARY;   PRT;   671 AA.
AC   Q84VE0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Cooper B., Hutchison D., Park S., Guimil S., Luginbuhl P., Ellero C.,
RA   Goff S., Glazebrook J.;
RT   "Identification of Rice (Oryza sativa) Proteins Linked to the Cyclin-
RT   Mediated Regulation of the Cell Cycle.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY224546; AAO72666.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.

```
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   671 AA;  72277 MW;  23F93B995CE21E60 CRC64;

  Query Match             28.6%;  Score 963;  DB 10;  Length 671;
  Best Local Similarity   37.5%;  Pred. No. 2.4e-67;
  Matches  253;  Conservative  118;  Mismatches  263;  Indels   40;  Gaps   19;

Qy         68 ELDGYSLTLGDV--VGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXN--SVYGVTTGF 123
Db         11 KIEGSSLRVGQVAAVSAAKDASGVVVELDEEARPRVKASSEWILNCIAHGGDIYGVTTGF 70

Qy        124 GGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSL 183
Db         71 GGTSHRRTKDGQALQVELLRHLNAGIFGNGSDG--------NSLPSEVSRAAMLVRINTL 122

Qy        184 TRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVX 243
Db        123 LQGYSGIRFEILEAITKLINTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQA-VT 181

Qy        244 HEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQ 301
Db        182 VDG--KKVDAAEAFKIAGIQGGFFRLEPKEGLAIVNGTSVGSALAAMVLYDANVLAVLSE 239

Qy        302 ALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDD 361
Db        240 VLSAVFCEVMNGKPEYTDHLTHKL-KHHPGSIEAAAIMEHILAGSAFMPHAQ---KVNEV 295

Qy        362 EGIL--RQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAG-QSTTDNPLIDVENKXTHHGG 418
Db        296 DPLLKPKQDRYALRTSPQWLGPQI-EVIRA-ATKSIEREVNSVNDNPVIDVHRGKALHGG 353

Qy        419 NFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGL 477
Db        354 NFQGTPIGVSMDNTRLAIANIGKLMFAQFSELVNEFYNNGLTSNLAGSRNPSLDYGFKGT 413

Qy        478 DIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYC 537
Db        414 EIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYLVA 473

Qy        538 VLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNS 597
Db        474 LCQAVDLRHLEENLKSAVKNCVTTVAKKVLTTGPAGGLHSARFSEKALLTAIDR-EAVYS 532

Qy        598 YDLEP--RWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAI--SLTRXVRXX 653
Db        533 YADDPCSANYPLMTKIRAVLVEHALANGPAEKDDGSSVFSKITAFEEELREALPREMEAA 592

Qy        654 FWXAPSSSSPALXYL-SPRTRVLYSFVREELGVKARRGDVFL-GKQEVTIGSNVSRIYEA 711
Db        593 RVAFETGTAPITNRIKESRSFPLYRFVREELGC------VYLTGEKLKSPGEECNKVFLA 646

Qy        712 IKSGRINXVLVKML 725
Db        647 ISERKLIDPMLECL 660

RESULT 5
Q8VXG7
ID   Q8VXG7     PRELIMINARY;      PRT;   703 AA.
AC   Q8VXG7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Corso;
RX   MEDLINE=97161122;
RA   Rosler J., Krekel F., Amrhein N., Schmid J.;
RT   "Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase
RT   activity.";
RL   Plant Physiol. 113:175-179(1997).
DR   EMBL; L77912; AAL40137.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   703 AA;  74926 MW;  11374FD68516971E CRC64;

  Query Match             28.5%;  Score 958.5;  DB 10;  Length 703;
  Best Local Similarity   34.6%;  Pred. No. 5.7e-67;
  Matches  255;  Conservative  127;  Mismatches  242;  Indels  113;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDV--VGAARKGRXVRVXDSDEI 97
Db         16 GAAAAELAGSHLDEVKRMVAQ-ARQPVVKIEGSTLRVGQVAAVASAKDASGVAVELDEEA 74

Qy         98 RXKIDKSVEFLRXQLXN--SVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXD 155
Db         75 RPRVKASSEWILDCIAHGGDIYGVTTGFGGTSHRRTKDGPALQVELLRHLNAGIFGTGSD 134

Qy        156 SFXLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGT 215
Db        135 G--------HTLPSEVTRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGT 186

Qy        216 ISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGL 273
Db        187 ITASGDLVPLSYIAGLITGRPNAQA-VTVDGRK--VDAAEAFKIAGIEGGFFKLNPKEGL 243

Qy        274 GLVNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQI 333
Db        244 AIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKL-KHHPGSI 302

Qy        334 EVARNIRTLLEGSXFAVHHEEEVKVKDDEGIL--RQDRYPLRTSPQWLGPLVSDMIHAHA 391
Db        303 EAAAIMEHILDGSSFM---KQAKKVNELDPLLKPKQDRYALRTSPQWLGPQI-EVIRA-A 357

Qy        392 VLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450
Db        358 TKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSEL 417

Qy        451 LNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509
Db        418 VNEFYNNGLTSNLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSADEHNQDV 477

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569
Db        478 NSLGLVSARKTAEAIDILKLMSSTYIVALCQAVDLRHLEENIK----------------- 520

Qy        570 XXTXXXXXXELXXKVXKXLXKRLEQTNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSL 629
Db        521 --------ASVKNTVTQVAKKVLTMNPSGEL--------SSARFSEKELISA-----IDR 559

Qy        630 AAVNAWKVAXAEKAISLTRXVRXXF----------------------WXAPSSSSP-- 663
Db        560 EAVFTYAEDAASASLPLMQKLRAVLVDHALSSGERGAGALRVLQDHQVRGGAPRGAAPGG 619

Qy        664 --------------ALXYLSPRTRVLYSFVREELGVKARRGDVFL-GKQEVTIGSNVSRI 708
```

```
Db        620 GGRPRGVAEGTAPVANRIADSRSFPLYRFVREELGC------VFLTGERLKSPGEECNKV 673

Qy        709 YEAIKSGRINXVLVKML 725
              :  |  |::   ::: |
Db        674 FVGISQGKLVDPMLECL 690

RESULT 6
Q9AXI5
ID   Q9AXI5      PRELIMINARY;      PRT;   711 AA.
AC   Q9AXI5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS   Pharbitis nil (Violet) (Japanese morning glory).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Solanales; Convolvulaceae; Ipomoea.
OX   NCBI_TaxID=35883;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Violet;
RA   Nakazawa A., Nozue M., Yasuda H., Takeba G., Kubo H.;
RT   "Expression of phenylalanine ammonia-lyase in Pharbitis nil.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF325496; AAG49585.1; -.
DR   HSSP; P21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   711 AA;  77085 MW;  CBE6E3A0C0969E90 CRC64;

  Query Match             28.4%;  Score 955;  DB 10;  Length 711;
  Best Local Similarity   36.3%;  Pred. No. 1.1e-66;
  Matches  259;  Conservative  103;  Mismatches  268;  Indels   84;  Gaps   18;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFL 108
              : || |:  :|:          |:||:  |     |  |:|  |  :  |  :::
Db         34 SHLDEVKVMVAE-FRKPAVKLGGETLTVAQVAAIASRDNAVTVELSEESRAGVKASSDWV 92

Qy        109 RXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLE-- 164
                  :        |||||||  ::   ||:    :|||  |:        |:          | | |
Db         93 MDSMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI---------FGNGTESC 143

Qy        165 NSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSP 224
              ::||    | || :|:|:| :|:| :|  :|||:|  |||  |||  :||||||:|||||  |
Db        144 HTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKLLNHNITPCLPLRGTITASGDLVP 203

Qy        225 LSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVS 282
              |||||  ||| |:||    |    |   :  :  |  ||:  |:          | |||||  |||||||
Db        204 LSYIAGLITGRPNSKA-VGPNG--ETLNAEEALRLAGVNGGFFELQPKEGLALVNGTAVG 260

Qy        283 ASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTL 342
              : ||::  |  :|::|::||: |:|:  | ||     |      |       :  ||  |||  |     :
Db        261 SGMASMVLFEANVLAVLSEVLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHI 319

Qy        343 LEGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTT 402
              |:|| :  |    :::    |       :||||  |||||||||  : ::|  |     :      |
Db        320 LDGSSY-VKAAQKMHEMDPLQKPKQDRYALRTSPQWLGPQI-EVIRAATKMIEREINSVN 377

Qy        403 DNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSC 462
              ||||||      |||||  :  :    :|: :||||  ||||  |  |  :|::|    |  ||||
Db        378 DNPLIDVARSKALHGGNFQGTPIGVSMDNSRLALASIGKLLFAQFSELVNDYYNNGLPSN 437

Qy        463 L-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTA 521
              |  |   :|||  |   ||  :||  |:|  |||   ||||| |||  ||    ||  ||||  |||||:||
Db        438 LTAGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISARKTA 497

Qy        522 EANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQ----------H----- 566
              || |||  |: :|:|   :  ||:|||  :|      :              :   :           |
Db        498 EAVDVLKLMSSTYLVALCQAIDLRFLEENLRNAVKNAVTQVAKRTLTMGANGELHPARFC 557

Qy        567 ------------FGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDLEPRWHDAFSXATG 613
                                  |                  |  |: :   |       :                  ||
Db        558 EKDLLRVVDREYVFAYADDPCSANYPLMQKLRQALVDHALQNGESEKN----------TG 607

Qy        614 TVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSSSSPAL--XYLSPR 671
              | :            : :||      |      | :    |              |   |  : ||:            |
Db        608 TSI---------FLKVAAFEDELKAVLPKEVEAAR-------IAVESGNPAIPNRIKECR 651

Qy        672 TRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
              :  || ||||  ||        ::   |::   :  |     ::: |:   |   |     |:: |
Db        652 SYPLYKFVREGLGT-----ELLTGEKVRSPGEECDKVFTAMCEGSIIDPLLECL 700

RESULT 7
Q9M567
ID   Q9M567      PRELIMINARY;      PRT;   730 AA.
AC   Q9M567;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
GN   PAL2.
OS   Rubus idaeus (Raspberry).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Rosales; Rosaceae; Rosoideae; Rubus.
OX   NCBI_TaxID=32247;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21437948; PubMed=11553751;
RA   Kumar A., Ellis B.E.;
RT   "The phenylalanine ammonia-lyase gene family in raspberry. structure,
RT   expression, and evolution.";
RL   Plant Physiol. 127:230-239(2001).
DR   EMBL; AF237955; AAF40224.1; -.
DR   HSSP; P21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   730 AA;  79422 MW;  0E942417E50CE113 CRC64;

  Query Match             28.4%;  Score 955;  DB 10;  Length 730;
  Best Local Similarity   36.8%;  Pred. No. 1.1e-66;
  Matches  257;  Conservative  107;  Mismatches  283;  Indels   52;  Gaps   19;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRX 99
              |:     :  || |    :|:           :|  |  :||:  |     |:    |:|   ::  |
Db         44 GAAAESMTGSHLDEVRRMVAE-YRKPVVKLGGETLTISQVAAIANHDSGVKVELAESARA 102

Qy        100 KIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSF 157
              :  |  :::     :            |||||||  ::   ||:   :|||  |:        |||     :|
Db        103 GVKASSDWVMDSMNKGTDSYGVTTGFGATSHRRTKQGAALQKELIRFLNAGVLRNGTES- 161
```

```
Qy        158 XLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTIS 217
                    ::||    | || :|:|:| :|:| :|   :|||:: |||| ||| :||||||:
Db        162 ------AHTLPHSATRAAMLVRINTLLQGYSGIRFEILEAISKFLNHNITPCLPLRGTIT 215

Qy        218 ASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGL 275
              ||||| ||||||  :|| |:||  |   :|   : : | || |   |:      | ||||| |
Db        216 ASGDLVPLSYIAGLLTGRPNSKA-VGPKG--ETLNAAEAFAQVGISSGFFELQPKEGLAL 272

Qy        276 VNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEV 335
              ||||| : :|: | : ::|:|||: |:|:  | | |     |     |     : || |||
Db        273 VNGTAVGSGLASTVLFETNILALLSEILSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEA 331

Qy        336 ARNIRTLLEGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSL 395
              |  :  :|:|| : |   |:: :|      :||||  ||||||||| :   :   :   |:
Db        332 AAIMEHILDGSSY-VKAAEKLHEQDPLQKPKQDRYALRTSPQWLGPQIE--VIRFSTKSI 388

Qy        396 EAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAG 454
              |      |  ||||||          ||||||  : :  :|: ||||:| |||  | | :|::|
Db        389 EREINSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLAIASIGKLMFAQFSELVNDF 448

Qy        455 MNRGLPSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLA 513
               | ||| |:    ||| |   || :|| |:| |||  |||||| ||| ||   || ||||
Db        449 YNNGLPSNLSGGRDPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLG 508

Qy        514 LISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTX 573
              |||:|:||| |:| |: :| |   : ||:||| :|    |           |   :
Db        509 LISSRKTAEAVDILKLMSSTFLVALCQAIDLRHLEENLKSTVKNTVSQLAKRVLTTGVNG 568

Qy        574 XXXXXELXXKVXKXLXKRLEQTNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVN 633
                          |     : :| |   :| |  :|        ||  :::  |           |  |  |
Db        569 ELHPSRFCEKDLLMVVER-EYLFAYIDDP------CSATYPLMQRLRQ---VLVEHALTN 618

Qy        634 A-----------WKVAXAEKAIS--LTRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSF 678
                          |:   |: :     | :      | |  : |:    :  |:  || |
Db        619 GENEKNASTSIFQKITAFEEELKTILPKEVESA-RAAYESGNAAIPNRIVECRSYPLYKF 677

Qy        679 VREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRI 717
              |||||      |:   |::  : |    ::: |:  |  |
Db        678 VREELG-----GEFLTGEKVRSPGEECDKVFTAMCQGNI 711
```

```
RESULT 8
Q94EN0
ID   Q94EN0      PRELIMINARY;      PRT;   708 AA.
AC   Q94EN0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   PAL1.
OS   Rehmannia glutinosa.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Lamiales; Lamiales incertae sedis; Rehmannia.
OX   NCBI_TaxID=99300;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Keumsan;
RA   Yun S.J., Lee B.K., Park M.R., Srinivas B., Chun J.-C.;
RT   "Molecular characterization of phenylalanine ammonia-lyase gene in
RT   Rhemannia glutinosa.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF401636; AAK84225.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   Lyase.
SQ   SEQUENCE   708 AA;  76835 MW;  1B6AF08816F1B8FE CRC64;

  Query Match             28.4%;  Score 954;  DB 10;  Length 708;
  Best Local Similarity   36.7%;  Pred. No. 1.3e-66;
  Matches  253;  Conservative  103;  Mismatches  260;  Indels   74;  Gaps   18;
```

```
Qy         68 ELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFGG 125
              :| | |||:  |    |  :    | |   ::   |  :  |  :::   :   |||||||
Db         50 KLGGESLTIAQVAAIAARDNAVAVELAETARAGVKASSDWVMESMNKGTDSYGVTTGFGA 109

Qy        126 SADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTR 185
              :: ||:  :||| |:     |:     :|          ::||    | || :|:|:| :
Db        110 TSHRRTKQGGALQKELIRFLNAGIFGNGTES-------NHALPHSATRAAMLVRINTLLQ 162

Qy        186 GHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHE 245
              |:| :|  :||||| |||| ||| :||||||:||||| |||||  :|| |:||  |
Db        163 GYSGIRFEILEALTKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPN 221

Qy        246 GXEKIMXAREAIALFGLEPVV-LGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALT 304
              |  :  : | || :| |:    | |||||| |||||||  : :|: | : ::|:|||: |:  :
Db        222 G--EALNAGEAFSLAGVSGFFELQPKEGLALVNGTAVGSGLASIALYDANILAVLSEVTS 279

Qy        305 ALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDDEGI 364
               :  |  |     |     :  || ||| |  :   :|:|| : |   :::   |
Db        280 VIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHETDPLQK 337

Qy        365 LRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQASA 424
               :|||| ||||||||| :  :   |  ::   |   |  | |||||      |||||| :
Db        338 PKQDRYALRTSPQWLGPQIEVIRTATKMIEREI-NSVNDTPLIDVSRNKALHGGNFQGTP 396

Qy        425 VXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAA 483
              :  :|: ||||:| |||| | |  |:|::|     |||| |:    :|||  |  :|| |:
Db        397 IGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGSEIAMAS 456

Qy        484 YTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQAVD 543
               | |||  |||||| ||| ||   || |||| |||:|:|: || |:| |: :|:|:  : ||||
Db        457 YCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTVEALDILKLMSSTYLIALCQAVD 516

Qy        544 LRAME----FEFKKQFXPXXXXLLXQHF-GXXXTXXXXXXELXXKVXKX---------- 587
              || :|       |          |        |            :|   | :
Db        517 LRHLEENLRLSVKNTVSQVAKRTLTMGINGELHPSRFCEKDLLRVVDREYVFAYIDDPCS 576

Qy        588 ----LXKRLEQT------NSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKV 637
                  | ::| |       | : |      |     :  :|         | | |:  :|
Db        577 GTYPLMQKLRQVLVDHALNNGESEKN----VSTSIFQKIEAFE------VELKAILPKEV 626

Qy        638 AXAEKAISLTRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSFVREELGVKARRGDVFLG 695
               |   |:                  | :||:        |:  || |:|||||        :   |
Db        627 ESARIAL--------------ESGNPAIGNRITECRSYPLYKFIREELGT-----NYLTG 667

Qy        696 KQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
              :: |: |    ::: |:  | |  |:| |
Db        668 EKVVSPGEECDKVFTALSKGLIVDPLLKCL 697
```

```
RESULT 9
Q8RUZ3
ID   Q8RUZ3      PRELIMINARY;      PRT;   681 AA.
AC   Q8RUZ3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5) (Fragment).
GN   PAL1.
OS   Pinus sylvestris (Scots pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3349;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Various strains;
RX   MEDLINE=21660210; PubMed=11801746;
RA   Dvornyk V., Sirvio A., Mikkonen M., Savolainen O.;
RT   "Low Nucleotide Diversity at the pal1 Locus in the Widely Distributed
RT   Pinus sylvestris.";
RL   Mol. Biol. Evol. 19:179-188(2002).
DR   EMBL; AF353967; AAL74317.1; -.
DR   EMBL; AF353968; AAL74318.1; -.
DR   EMBL; AF353969; AAL74319.1; -.
DR   EMBL; AF353970; AAL74320.1; -.
DR   EMBL; AF353971; AAL74321.1; -.
DR   EMBL; AF353972; AAL74322.1; -.
DR   EMBL; AF353973; AAL74323.1; -.
DR   EMBL; AF353976; AAL74326.1; -.
DR   EMBL; AF353978; AAL74328.1; -.
DR   EMBL; AF353979; AAL74329.1; -.
DR   EMBL; AF353980; AAL74330.1; -.
DR   EMBL; AF353982; AAL74332.1; -.
DR   EMBL; AF353983; AAL74333.1; -.
DR   EMBL; AF353984; AAL74334.1; -.
DR   EMBL; AF353985; AAL74335.1; -.
DR   EMBL; AF353986; AAL74336.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
FT   NON_TER       1      1
FT   NON_TER     681    681
SQ   SEQUENCE   681 AA;  74365 MW;  5B4521F7978BF804 CRC64;

  Query Match            28.1%;  Score 944.5;  DB 10;  Length 681;
  Best Local Similarity   43.9%;  Pred. No. 7e-66;
  Matches  214;  Conservative   85;  Mismatches  162;  Indels   27;  Gaps   10;

Qy         69 LDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLX--NSVYGVTTGFGGS 126
              ::| |||: ||   ||: :     |::  : ::::|  ::  |:      ||||||| :
Db         56 IEGKSLTISDVAAVARRSQVKVKLDAEAAKSRVEESSNWVLTQMTKGTDTYGVTTGFGAT 115

Qy        127 ADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTRG 186
               :  ||      ||| |:      ||            ||:  || |   :   | ||  :| |:| :|
Db        116 SHRRTNQGAELQKELIRFLNAGV---------LGKCPENVLSEDTTRAAMLVRTNTLLQG 166

Qy        187 HSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEG 246
              :| :|  :|| :    || |:|| :|||||:|||||  ||||||  :|| |:|:|     :|
Db        167 YSGIRWDILETVEKLLNAGLTPKLPLRGTITASGDLVPLSYIAGLLTGRPNSRVR-SRDG 225

Qy        247 XEKIMXAREAIALFGLE-PVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTA 305
                 |  |   ||:   |||  |   |  |||||  :||||:|  |::|::     ||::|:|||: ::|
Db        226 IE--MSGAEALKKVGLEKPFELQPKEGLAIVNGTSVGAALASIVCFDANVLALLSEVISA 283

Qy        306 LTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH----HEEEVKVKDD 361
               :  |  |  |        |   |  : :  || |:|  |   :  :|:||  :   |     ||      |
Db        284 MFCEVMNGKPEFTDPLTHKL-KHHPGQMEAAAIMEYVLDGSSYMKHAAKLHEMNPLQKP- 341

Qy        362 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQ 421
                  :||||  ||||||||||  |   :    |    ::    |      |    |||:|||      |||||
Db        342 ----KQDRYALRTSPQWLGPQVEIIRSATHMIEREI-NSVNDNPVIDVARDKALHGGNFQ 396

Qy        422 ASAVXNTMEXTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIA 480
               :   :   :|:   ||:::  ||||  |  |  :|::|      |||||  |:    :|||   |   ||:||
Db        397 GTPIGVSMDNLRLSISAIGKLMFAQFSELVNDYYNGGLPSNLSGGPNPSLDYGLKGAEIA 456

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
               |:|||||  :||||||:|||  ||     ||  ||||  |:|||::|||  |:|  |:|:|:|  :  |
Db        457 MASYTSELLYLANPVTSHVQSAEQHNQDVNSLGLVSARKSAEAIDILKLMLSTYLTALCQ 516

Qy        541 AVDLRAME 548
              ||||| :|
Db        517 AVDLRHLE 524


RESULT 10
Q852S1
ID   Q852S1      PRELIMINARY;      PRT;   715 AA.
AC   Q852S1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN   GDCPAL3.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC   Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Kurodagosun;
RA   Ozeki Y., Chikagawa Y., Kimura S., Soh H., Maeda K., Pornsiriwong W.,
RA   Kato M., Akimoto H., Oyanagi M., Fukuda T., Koda T., Itoh Y.,
RA   Yamada A., Ueno H., Takeda J.;
RT   "Putative cis-elements in the promoter region of phenylalanine
RT   ammonia-lyase gene of carrot involved in anthocyanin biosynthetic
RT   pathway.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB089813; BAC56977.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   715 AA;  77802 MW;  8D067795D57565B4 CRC64;

  Query Match            28.1%;  Score 944.5;  DB 10;  Length 715;
  Best Local Similarity   36.0%;  Pred. No. 7.5e-66;
  Matches  251;  Conservative  111;  Mismatches  250;  Indels   85;  Gaps   19;

Qy         68 ELDGYSLTLGDVVG-AARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFG 124
              :|  | :||:  |    :||    |:  |:  |    :    | :::    :        |||||||
Db         54 KLGGETLTISQVAAISARDDSGVKVELSEAARAGVKASSDWVMESMNKGTDSYGVTTGFG 113

Qy        125 GSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLE---NSLPLEVVRGAMTIRVN 181
               ::  ||:   :|||  |:      |:          | |  |    |:||    |  || :|:|
Db        114 ATSHRRTKQGGALQKELIRFLNAGI---------FGSGAEAGNNTLPHSATRAAMLVRIN 164

Qy        182 SLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVH 241
              :|  :|:|  :|   :|||:|  ||||  |||  :||||||:|||||  ||||||    :||  |:||
Db        165 TLLQGYSGIRFEILEAITKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA- 223

Qy        242 VXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLL 299
               |     |     :     ||   |  |:|      |||||  ||||||  :  ||::   :|::|::|
Db        224 VGPTGV--TLSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVL 281

Qy        300 SQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVK 359
               :: ::|:   |   |      |     |   :  ||  |||  |   :   :|:||  :  |      :::
Db        282 AEVMSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEM 339
```

```
Qy        360 DDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGN 419
Db        340 DPLQKPKQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGN 398

Qy        420 FQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLD 478
Db        399 FQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAE 458

Qy        479 IAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCV 538
Db        459 IAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGL 518

Qy        539 LQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKV------------- 584
Db        519 CQAIDLRHLEENLK-----------------STVKNTVSQVAKRVLTMGVNGELHPSRF 560

Qy        585 -XKXLXKRLEQ----------TNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAV 632
Db        561 CEKDLLRVVDREYIFAYIDDPCSATYPLMQKLRE-------TLVEHALNNGDKERNLSTS 613

Qy        633 NAWKVAXAEKAIS--LTRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSFVREELGVKAR 688
Db        614 IFQKIAAFEDELKALLPKEVESA-RAAVESGNPAIPNRIKECRSYPLYKFVREELGT--- 669

Qy        689 RGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
Db        670 --EYLTGEKVTSPGEEFDKVFTAMTKGEIIDPLLECL 704
```

RESULT 11
Q8SAS6

```
ID   Q8SAS6      PRELIMINARY;      PRT;   681 AA.
AC   Q8SAS6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5) (Fragment).
GN   PAL1.
OS   Pinus sylvestris (Scots pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3349;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kir30;
RX   MEDLINE=21660210; PubMed=11801746;
RA   Dvornyk V., Sirvio A., Mikkonen M., Savolainen O.;
RT   "Low Nucleotide Diversity at the pal1 Locus in the Widely Distributed
RT   Pinus sylvestris.";
RL   Mol. Biol. Evol. 19:179-188(2002).
DR   EMBL; AF353981; AAL74331.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
FT   NON_TER       1      1
FT   NON_TER     681    681
SQ   SEQUENCE   681 AA;  74341 MW;  5BFCE337969FFBD0 CRC64;

  Query Match             28.0%;  Score 943.5;  DB 10;  Length 681;
  Best Local Similarity   43.9%;  Pred. No. 8.4e-66;
  Matches  214;  Conservative   85;  Mismatches  162;  Indels   27;  Gaps   10;
```

```
Qy         69 LDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLX--NSVYGVTTGFGGS 126
Db         56 IEGKSLTISDVAAVARRSQVKVKLDAEAAKSRVEESSNWVLTQMTKGTDTYGVTTGFGAT 115

Qy        127 ADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTRG 186
Db        116 SHRRTNQGAELQKELIRFLNAGV---------LGKCPENVLSEDTTRAAMLVRTNTLLQG 166

Qy        187 HSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEG 246
Db        167 YSGIRWDILETVEKLLNAGLTPKLPLRGTITASGDLVPLSYIAGLLTGRPNSRVR-SRDG 225

Qy        247 XEKIMXAREAIALFGLE-PVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTA 305
Db        226 IE--MSGAEALKKVGLEKPFELQPKEGLAIVNGTSVGAALASIVCFDANVLALLSEVISA 283

Qy        306 LTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH----HEEEVKVKDD 361
Db        284 MFCEVMNGKPEFTDPLTHKL-KLHPGQMEAAAIMEYVLDGSSYMKHAAKLHEMNPLQKP- 341

Qy        362 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQ 421
Db        342 ----KQDRYALRTSPQWLGPQVEIIRSATHMIEREI-NSVNDNPVIDVARDKALHGGNFQ 396

Qy        422 ASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIA 480
Db        397 GTPIGVSMDNLRLSISAIGKLMFAQFSELVNDYYNGGLPSNLSGGPNPSLDYGLKGAEIA 456

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        457 MASYTSELLYLANPVTSHVQSAEQHNQDVNSLGLVSARKSAEAIDILKLMLSTYLTALCQ 516

Qy        541 AVDLRAME 548
Db        517 AVDLRHLE 524
```

RESULT 12
Q9MAX1

```
ID   Q9MAX1      PRELIMINARY;      PRT;   716 AA.
AC   Q9MAX1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia lyase.
OS   Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   lamiids; Gentianales; Apocynaceae; Rauvolfioideae; Vinceae;
OC   Catharanthus.
OX   NCBI_TaxID=4058;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kiyota S., Sakano K., Kim J.;
RT   "cDNA of phenylalanine ammmonia lyase from Catharanthus roseus.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB042520; BAA95629.1; -.
DR   HSSP; P21310; 1B8F.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   716 AA;  78241 MW;  1F16D4DEB8683B0B CRC64;
```

```
  Query Match             28.0%;  Score 942.5;  DB 10;  Length 716;
  Best Local Similarity   35.7%;  Pred. No. 1.1e-65;
  Matches  256;  Conservative  120;  Mismatches  270;  Indels   71;  Gaps   22;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVG-AARKGRXVRVXDSDEIR 98
              |    :    : || |:  :|:           :|  | :||:  |   |||    |:|  |:: |
Db         29 GMAADSLKGSHLDEVKRMVAE-FRKPVVKLGGETLTISQVAAIAARDYNAVKVELSEDAR 87

Qy         99 XKIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDS 156
                :  | :::   :         ||||||| ::  ||:   :||| |:      |:
Db         88 AGVKASSDWVMESMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI------- 140

Qy        157 FXLGRGLENS--LPLEVVRGAMTIRVNSLTRGHSAVRLVLEALTNFLNHGITPIVPLRG 214
                | | |:|  ||         | || :|:|:| :|:| :|  :|||:| |||| ||| :||||
Db        141 --FGNGTESSHTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNHNITPCLPLRG 198

Qy        215 TISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEPVV--LGPKEG 272
              ||:|||| |||||  :|| |:||   |     |   :|:   :|   :  |:    :   ||||
Db        199 TITASGDLVPLSYIAGLLTGRPNSKA-VGPNG--EIVNPEQAFKMAGVNDGLFELQPKEG 255

Qy        273 LGLVNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQ 332
              | ||||||  : ||::  | :|::|::|| |:|:  | | |      |      |    : || |
Db        256 LALVNGTAVGSGMASMVLFEANILAVLSEVLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQ 314

Qy        333 IEVARNIRTLLEGSXF--AVH--HEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIH 388
              || |   :  :|:|| :   |  |  || :   |        :|||  ||||||||| :   :
Db        315 IEAAAIMEHILDGSGYVKAAHKLHEMDPLQKP-----KQDRYALRTSPQWLGPQIEVIRS 369

Qy        389 AHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLT 448
              |  ::   |    | |||||||      |||||  :  :  :|:      |||:|  ||| |  |::
Db        370 ATKMIEREI-NSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNARLAIASIGKLMFAQIS 428

Qy        449 EMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQ 507
              |::|     | ||| |:   :||| |   || :|| |:| ||| :||||| ||| || ||
Db        429 ELVNDFYNNGLPSNLSGGRNPSLVYGFKGAEIAMASYCSELQYLANPVTNHVQSAEQHNQ 488

Qy        508 AVNSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHF 567
               |||| |||:|:|||| ::| |: :|:|   : ||:||| :|    |              :   :
Db        489 DVNSLGLISSRKTAEAVEILKLMSSTYLVALCQAIDLRHLEENLKNAVKNTVSQVAKRTL 548

Qy        568 -----GXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDLEPRWHDAFSXATGT-------- 614
                         |          |      :         :|            :||
Db        549 TVGVNGELHPSRFCEKDLIRVVDR------EYVFAYVDDP--------CSGTYPLMEKLR 594

Qy        615 --VVEXLSSXXXXXVSLAAVNAWKVAXAEKAIS--LTRXVRXXFWXAPSSSSPAL--XYL 668
                 :|:           ::          |:|   |     :   |  :|   |  |     | : :||:
Db        595 QVIVDHALQNGESEKNVNTSIFQKIAAFEDELKTVLPKEVESA-RTALENGNPAIPNRIK 653

Qy        669 SPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
                  |:  || ||||::|          :     |::: :  |      ::: |: : :|    |:: |
Db        654 ECRSYPLYKFVREDVG-----AEFLTGEKDRSPGEEFDKVFTAMCNEKIIDPLLECL 705

RESULT 13
Q8RV49
ID   Q8RV49       PRELIMINARY;      PRT;    681 AA.
AC   Q8RV49;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase (EC 4.3.1.5) (Fragment).
GN   PAL1.
OS   Pinus sylvestris (Scots pine).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3349;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kol3, and Kol4;
RX   MEDLINE=21660210; PubMed=11801746;
RA   Dvornyk V., Sirvio A., Mikkonen M., Savolainen O.;
RT   "Low Nucleotide Diversity at the pal1 Locus in the Widely Distributed
RT   Pinus sylvestris.";
```

```
RL   Mol. Biol. Evol. 19:179-188(2002).
DR   EMBL; AF353974; AAL74324.1; -.
DR   EMBL; AF353975; AAL74325.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
FT   NON_TER       1      1
FT   NON_TER     681    681
SQ   SEQUENCE   681 AA;  74394 MW;  35E8A50BA8480286 CRC64;

  Query Match             28.0%;  Score 941.5;  DB 10;  Length 681;
  Best Local Similarity   43.9%;  Pred. No. 1.2e-65;
  Matches  214;  Conservative   84;  Mismatches  163;  Indels   27;  Gaps   10;

Qy         69 LDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLX--NSVYGVTTGFGGS 126
              ::| |||: ||    ||: :         | :   : ::::|  ::   |:           ||||||| :
Db         56 IEGKSLTISDVAAVARRSQVKVKLDVEAAKSRVEESSNWVLTQMTKGTDTYGVTTGFGAT 115

Qy        127 ADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTRG 186
               :  ||      ||| |:     ||              ||:  || |    :   | || :| |:| :|
Db        116 SHRRTNQGAELQKELIRFLNAGV---------LGKCPENVLSEDTTRAAMLVRTNTLLQG 166

Qy        187 HSAVRLVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEG 246
              :| :|  :|| :    || |:|| :||||||:||||| ||||||    :|| |:|:|      :|
Db        167 YSGIRWDILETVEKLLNAGLTPKLPLRGTITASGDLVPLSYIAGLLTGRPNSRVR-SRDG 225

Qy        247 XEKIMXAREAIALFGLE-PVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTA 305
               |  |   ||:    |||  |   | |||||  :||||:|  |::|::     ||::|:|||: ::|
Db        226 IE--MSGAEALKKVGLEKPFELQPKEGLAIVNGTSVGAALASIVCFDANVLALLSEVISA 283

Qy        306 LTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH----HEEEVKVKDD 361
               :  | |  |        | :  :| |:|  |    :  :|:|:|  :     |            ||
Db        284 MFCEVMNGKPEFTDPLTHKL-KHHPGQMEAAAIMEYVLDGSSYMKHAAKLHEMNPLQKP- 341

Qy        362 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQ 421
                  :|||| |||||||||| |    :   |  ::    |      |||:|||          ||||||
Db        342 ----KQDRYALRTSPQWLGPQVEIIRSATHMIEREI-NSVNDNPVIDVARDKALHGGNFQ 396

Qy        422 ASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIA 480
               :  :  :|:   ||:::: ||||  |   |  :|::|     | ||||  |:      :|||  |     || :||
Db        397 GTPIGVSMDNLRLSISAIGKLMFAQFSELVNDYYNGGLPSNLSGGPNPSLDYGLKGAEIA 456

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
               |:||||| :||||||:|||  ||    || ||||||·|:|||::|||  |:| |:|:|:|   :  |
Db        457 MASYTSELLYLANPVTSHVQSAEQHNQDVNSLGLVSARKSAEAIDILKLMLSTYLTALCQ 516

Qy        541 AVDLRAME 548
              |||||  :|
Db        517 AVDLRHLE 524

RESULT 14
Q8RWP4
ID   Q8RWP4       PRELIMINARY;      PRT;    717 AA.
AC   Q8RWP4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase.
GN   AT3G53260.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
```

```
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY092957; AAM12956.1; -.
DR   EMBL; BT000035; AAN15354.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   717 AA;  77801 MW;  C4C9CBDCDB91A838 CRC64;

  Query Match             27.9%;  Score 940;  DB 10;  Length 717;
  Best Local Similarity   35.1%;  Pred. No. 1.7e-65;
  Matches  250;  Conservative  122;  Mismatches  259;  Indels   82;  Gaps   20;

Qy         49 TQLDIVEXXLAD---PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSV 105
Db         40 SHLDEVKKMVAEYRRPVVN----LGGETLTIGQVAAISTVGGSVKVELAETSRAGVKASS 95

Qy        106 EFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGL 163
Db         96 DWVMESMNKGTDSYGVTTGFGATSHRRTKNGTALQTELIRFLNAGIFGNTKETC------ 149

Qy        164 ENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLS 223
Db        150 -HTLPQSATRAAMLVRVNTLLQGYSGIRFEILEAITSLLNHNISPSLPLRGTITASGDLV 208

Qy        224 PLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAV 281
Db        209 PLSYIAGLLTGRPNSKA-TGPDG--ESLTAKEAFEKAGISTGFFDLQPKEGLALVNGTAV 265

Qy        282 SASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRT 341
Db        266 GSGMASMVLFEANVQAVLAEVLSAIFAEVMSGKP-EFTDHLTHRLKHHPGQIEAAAIMEH 324

Qy        342 LLEGSXF-----AVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLE 396
Db        325 ILDGSSYMKLAQKVHEMDPLQKP------KQDRYALRTSPQWLGPQIE--VIRQATKSIE 376

Qy        397 AG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGM 455
Db        377 REINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFY 436

Qy        456 NRGLPSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLAL 514
Db        437 NNGLPSNLTASSNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGL 496

Qy        515 ISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXX 574
Db        497 ISSRKTSEAVDILKLMSTTFLVGICQAVDLRHLEENLRQ------------------TVK 538

Qy        575 XXXXELXXKV--------------XKXLXKRL--EQTNSYDLEP--RWHDAFSXATGTV 615
Db        539 NTVSQVAKKVLTTGINGELHPSRFCEKDLLKVVDREQVFTYVDDPCSATYPLMQRLRQVI 598

Qy        616 VEXLSSXXXXXVSLAAVNAWKVAXAEKAIS--LTRXVRXXFWXAPSSSSPALXYLSP-RT 672
Db        599 VDHALSNGETEKNAVTSIFQKIGAFEEELKAVLPKEVEAARAAYGNGTAPIPNRIKECRS 658

Qy        673 RVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
Db        659 YPLYRFVREELGTK-----LLTGEKVVSPGEEFDKVFTAMCEGKLIDPLMDCL 706

RESULT 15
Q8GZR8
ID   Q8GZR8      PRELIMINARY;      PRT;   713 AA.
AC   Q8GZR8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phenylalanine ammonia-lyase2.
OS   Lactuca sativa (Garden lettuce).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;
OC   Lactuca.
OX   NCBI_TaxID=4236;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Campos-Vargas R., Nonogaki H., Suslow T., Saltveit M.E.;
RT   "Characterization of a second phenylalanine ammonia-lyase (PAL) gene
RT   in wounded lettuce leaf tissue.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF411134; AAO13347.1; -.
DR   GO; GO:0005737; C:cytoplasm; IEA.
DR   GO; GO:0016211; F:ammonia ligase activity; IEA.
DR   GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR   GO; GO:0016829; F:lyase activity; IEA.
DR   GO; GO:0009058; P:biosynthesis; IEA.
DR   GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR   InterPro; IPR008948; L-Aspartase-like.
DR   InterPro; IPR001106; Phe/His_NH3lyase.
DR   InterPro; IPR005922; Phe_am_lyase.
DR   Pfam; PF00221; PAL; 1.
DR   TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR   PROSITE; PS00488; PAL_HISTIDASE; 1.
KW   Lyase.
SQ   SEQUENCE   713 AA;  77657 MW;  87751A2ED7349E61 CRC64;

  Query Match             27.9%;  Score 939;  DB 10;  Length 713;
  Best Local Similarity   36.1%;  Pred. No. 2e-65;
  Matches  258;  Conservative  106;  Mismatches  265;  Indels   86;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVG-AARKGRXVRVXDSDEIR 98
Db         29 GMAAESLKGSHLDEVKRMVAE-FRKPVVRLGGETLTVSQVAAIAASDNAGVKVELSETAR 87

Qy         99 XKIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDS 156
Db         88 AGVKASSDWVMESMNKGTDSYGVTTGFGATSHRRTKEGGALQKELIRFLNAGI------- 140

Qy        157 FXLGRGLE--NSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRG 214
Db        141 --FGNGTESTHTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNHNVTPFLPLRG 198

Qy        215 TISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEG 272
Db        199 TITASGDLVPLSYIAGLLTGRANSKA-VGPTG--EVLNAEKAFAEAGVEGGFFELQPKEG 255
```

```
Qy        273 LGLVNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQ 332
              | |||||||  :  ||::  | ||::|:|||:  |:|:   | | |    |    |     : || |
Db        256 LALVNGTAVGSGMASMVLFDANVLALLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQ 314

Qy        333 IEVARNIRTLLEGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAV 392
              || |  :  :|:|| :  |    ::|    |      :|||| |||||||||| : ::| :
Db        315 IEAAAIMEYILDGSDY-VKAAQKVHEMDPLQKPKQDRYALRTSPQWLGPQI-EVIRSSTK 372

Qy        393 LSLEAGQSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLN 452
              :         |   ||||||     |||||| : :   :|: ||||:| ||||  |  | :|::|
Db        373 MIEREINSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVN 432

Qy        453 AGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNS 511
                  | |||| |:    :|||  |  || :||  |:|  |||   |||||  |||  ||   || |||
Db        433 DFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNS 492

Qy        512 LALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHF---- 567
              | |||||:|||: ::| |:  |:|  : |::|||  :| |  |          |: :
Db        493 LGLISARKTAESVEILKLMSTTYLVALCQSIDLRHLE-ENLKSTVKNTVSLVAKKILTTG 551

Qy        568 --GXXXTXXXXXXELXXKVXKX--------------LXKRLEQ-------TNSYDLEPR 603
                |             :|   |  :                  | ::| |           |  |
Db        552 VNGELHPSRFCEKDLLRVVDREYVFAYIDDACSATYPLMQKLRQVIVDHALNNEND---- 607

Qy        604 WHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSSSSP 663
                     |  :: : :|        ::          |  | :: :||              |:
Db        608 -------AGTSIFQKISEFEEELKAVLPKEVEGVRSAYESSTLT---------IPNRIKE 651

Qy        664 ALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTI-GSNVSRIYEAIKSGRI 717
                |          || |||||||       ||  :|||  |     ::: |:  | |
Db        652 CRSY------PLYRFVREELGTG------FLTGEEVTSPGEEFDKVFTALCKGHI 694
```

Search completed: September 9, 2004, 10:07:45
Job time : 91.3651 secs

GenCore version 5.1.6


```
OM protein - protein search, using sw model

Run on:         September  9, 2004, 09:49:43 ; Search time 89.3693 Seconds
                                          (without alignments)
                                          2295.302 Million cell updates/sec

Title:          US-09-939-408A-21
Perfect score:  3365
Sequence:       1 MAPSLDSIATSXANGXXNGX..........RIYEAIKSGRINXVLVKMLA 726

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
                1:  geneseqp1980s:*
                2:  geneseqp1990s:*
                3:  geneseqp2000s:*
                4:  geneseqp2001s:*
                5:  geneseqp2002s:*
                6:  geneseqp2003as:*
                7:  geneseqp2003bs:*
                8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3223	95.8	726	5	ABB07694	Abb07694 Rhodotoru
2	3223	95.8	726	5	AAE27942	Aae27942 Phenylala
3	3021	89.8	716	1	AAP80513	Aap80513 L-phenyla
4	3021	89.8	716	5	ABB07693	Abb07693 R. torulo
5	3021	89.8	716	5	AAE27941	Aae27941 Rhodotoru
6	3021	89.8	716	5	AAE20653	Aae20653 Rhodotoru
7	3021	89.8	716	5	AAE16389	Aae16389 Rhodotoru
8	3021	89.8	716	6	ABG71948	Abg71948 R. glutin
9	3021	89.8	716	7	ADE36054	Ade36054 Rhodospor
10	3019	89.7	716	5	AAE20668	Aae20668 R. glutin
11	3019	89.7	716	5	AAE20654	Aae20654 Rhodotoru
12	3019	89.7	716	5	AAE16390	Aae16390 Rhodotoru
13	3019	89.7	716	6	ABG71955	Abg71955 R. glutin
14	3019	89.7	716	7	ADE36056	Ade36056 Rhodospor
15	3018	89.7	716	1	AAP81099	Aap81099 Sequence
16	3016	89.6	716	7	ADE36076	Ade36076 Rhodospor
17	3015	89.6	716	1	AAP83141	Aap83141 Sequence
18	3015	89.6	716	7	ADE36074	Ade36074 Rhodospor
19	3014	89.6	716	5	AAE20663	Aae20663 Rhodotoru
20	3014	89.6	716	6	ABG71954	Abg71954 R. glutin
21	3014	89.6	716	7	ADE36068	Ade36068 Rhodospor
22	3013	89.5	716	5	AAE20667	Aae20667 R. glutin
23	3013	89.5	716	7	ADE36060	Ade36060 Rhodospor
24	3012	89.5	716	5	AAE20665	Aae20665 R. glutin
25	3012	89.5	716	5	AAE20711	Aae20711 R. glutin
26	3012	89.5	716	5	AAE20660	Aae20660 Rhodotoru
27	3012	89.5	716	6	ABG71951	Abg71951 R. glutin
28	3012	89.5	716	7	ADE36062	Ade36062 Rhodospor
29	3011	89.5	716	7	ADE36066	Ade36066 Rhodospor
30	3010	89.5	716	5	AAE20662	Aae20662 Rhodotoru
31	3010	89.5	716	5	AAE20661	Aae20661 Rhodotoru
32	3010	89.5	716	5	AAE20666	Aae20666 R. glutin
33	3010	89.5	716	6	ABG71952	Abg71952 R. glutin
34	3010	89.5	716	6	ABG71953	Abg71953 R. glutin
35	3007	89.4	716	1	AAP95783	Aap95783 L-phenyla
36	3006	89.3	716	5	AAE20659	Aae20659 Rhodotoru
37	3006	89.3	716	6	ABG71950	Abg71950 R. glutin
38	3006	89.3	716	7	ADE36070	Ade36070 Rhodospor
39	3005	89.3	716	5	AAE20664	Aae20664 R. glutin
40	3005	89.3	716	7	ADE36058	Ade36058 Rhodospor
41	3005	89.3	716	7	ADE36072	Ade36072 Rhodospor
42	2997	89.1	716	7	ADE36064	Ade36064 Rhodospor
43	2978	88.5	716	5	AAE20658	Aae20658 Rhodotoru
44	2978	88.5	716	6	ABG71949	Abg71949 R. glutin
45	2971	88.3	686	5	AAE20657	Aae20657 Rhodotoru

ALIGNMENTS

```
RESULT 1
ABB07694
ID   ABB07694 standard; protein; 726 AA.
XX
AC   ABB07694;
XX
DT   29-AUG-2003  (revised)
DT   07-AUG-2003  (revised)
DT   10-JUN-2002  (first entry)
XX
DE   Rhodotorula PAL consensus polypeptide sequence.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
KW   consensus; enzyme.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1. .726
FT                   /note= "Xaa is any amino acid; "Xaa" indicates no
FT                   consensus at that position"
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   N-PSDB; ABA95244.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Claim 10; Fig 2; 135pp; English.
XX
```

```
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a consensus polypeptide sequence of the Rhodotorula PAL.
CC   (Updated on 07-AUG-2003 to correct OS field.) (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 726 AA;

  Query Match             95.8%;  Score 3223;  DB 5;  Length 726;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  726;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        721 LVKMLA 726
              ||||||
Db        721 LVKMLA 726
```

```
RESULT 2
AAE27942
ID   AAE27942 standard; protein; 726 AA.
XX
AC   AAE27942;
XX
DT   06-AUG-2003  (revised)
DT   27-DEC-2002  (first entry)
XX
DE   Phenylalanine ammonia lyase consensus protein.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide.
XX
OS   Rhodotorula graminis.
OS   Rhodotorula mucilaginosa.
OS   Rhodosporidium toruloides.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 726. .719
FT                   /note= "Xaa corresponds to unknown amino acid residue"
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Claim 26; Page 47-48; 74pp; English.
XX
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is PAL consensus protein.
CC   (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 726 AA;

  Query Match             95.8%;  Score 3223;  DB 5;  Length 726;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  726;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
```

```
Db        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Qy        721 LVKMLA 726
Db        721 LVKMLA 726

RESULT 3
AAP80513
ID   AAP80513 standard; protein; 716 AA.
XX
AC   AAP80513;
XX
DT   25-MAR-2003  (revised)
DT   12-NOV-1990  (first entry)
XX
DE   L-phenylalanine ammonialyase.
XX
KW   L-phenylalanine ammonialyase; PAL; foreign gene expression;
KW   culture temperature; expression regulation.
XX
OS   Rhodosporidium toruloides.
XX
PN   EP279665-A.
XX
PD   24-AUG-1988.
XX
PF   18-FEB-1988;   88EP-00301356.
XX
PR   19-FEB-1987;   87JP-00034397.
PR   18-JUN-1987;   87JP-00152359.
XX
PA   (MITK ) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N,  Yoshino S,  Yamamoto K,  Sone S,  Suzuki M,  Nakajima Y;
XX
DR   WPI; 1988-236895/34.
DR   N-PSDB; AAN81116.
XX
PT   Regulation of expression of foreign gene in Escherichia coli - by
PT   maintaining temp. of culture at 40 deg. C or more to suppress expression.
XX
PS   Claim 3; Page 15-19; 36pp; English.
XX
CC   The PAL gene is expressed in E.coli carrying a recombinant plasmid.
CC   Expression is regulated by maintaining the temperature at at least 40
CC   deg. C to suppress expression. Pal is produced in high concentrations.
CC   See also EP-279664: regualtion by sugar component as C source at 0.3% or
CC   more. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
CC   2003 to correct PI field.)
XX
SQ   Sequence 716 AA;

  Query Match              89.8%;  Score 3021;  DB 1;  Length 716;
  Best Local Similarity    84.0%;  Pred. No. 4.5e-300;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716

RESULT 4
ABB07693
ID   ABB07693 standard; protein; 716 AA.
```

```
XX
AC   ABB07693;
XX
DT   07-AUG-2003  (revised)
DT   10-JUN-2002  (first entry)
XX
DE   R. toruloides PAL polypeptide.
XX
KW   PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; enzyme;
KW   cytostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
OS   Rhodosporidium toruloides.
XX
PN   WO200208402-A2.
XX
PD   31-JAN-2002.
XX
PF   24-JUL-2001; 2001WO-US023270.
XX
PR   24-JUL-2000; 2000US-00624693.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-268973/31.
DR   N-PSDB; ABA95243.
XX
PT   Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT   treating mammal having disease or disorder from phenylketonuria, cancer,
PT   human immunodeficiency virus infection and human cytomegalovirus
PT   infection.
XX
PS   Example 3; Fig 1; 135pp; English.
XX
CC   The invention relates to an isolated and purified yeast phenylalanine
CC   ammonia-lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC   host cell comprising the polynucleotide is useful for the production of L
CC   -PAL, by adding the polynucleotide or the host cell to a composition
CC   comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC   improving the production of PAL, its analog or another optically active
CC   unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC   are useful for treating a mammal having a disease, disorder or condition
CC   selected from phenylketonuria, cancer, human immunodeficiency virus
CC   infection and human cytomegalovirus infection. The present sequence
CC   represents a R. toruloides PAL polypeptide. (Updated on 07-AUG-2003 to
CC   correct OS field.)
XX
SQ   Sequence 716 AA;

  Query Match             89.8%;  Score 3021;  DB 5;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 4.5e-300;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||           | |||  ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   |||||| ||||| |||||| ||| |||||| ||||||||| || |||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||||:||||||::|||||||||||||||||| :|||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| |||||||||||||||| |:|||||||||||||||||
Db        474 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:|||||||| |      |: ||||        |     ||  || | | ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               ||||||| | ||||| |||      :|||||||||| || |||||| || | | :
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| |||||::||:||||||||||||||||||||||||||||||:|||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716


RESULT 5
AAE27941
ID   AAE27941 standard; protein; 716 AA.
XX
AC   AAE27941;
XX
DT   06-AUG-2003  (revised)
DT   27-DEC-2002  (first entry)
XX
DE   Rhodotorula toruloides PAL protein.
XX
KW   Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW   cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW   human cytomegalovirus infection; cytostatic; virucide.
XX
OS   Rhodosporidium toruloides.
XX
PN   US2002102712-A1.
XX
PD   01-AUG-2002.
XX
PF   24-AUG-2001; 2001US-00939408.
XX
PR   24-JUL-2000; 2000US-00624693.
PR   24-JUL-2001; 2001WO-US023270.
XX
PA   (PCBU-) PCBU SERVICES INC.
XX
PI   Yoshida RK,  Kootstra AB;
XX
DR   WPI; 2002-690616/74.
DR   N-PSDB; AAD45815.
XX
PT   Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT   polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT   human immunodeficiency virus or human cytomegalovirus infection.
XX
PS   Example 3; Page 43-45; 74pp; English.
XX
```

```
CC   The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC   ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC   such proteins. PAL sequences are useful for producing L-phenylalanine.
CC   They are useful for treating mammals having diseases, disorders or
CC   conditions that would benefit from treatment with PAL proteins such as
CC   phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC   human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC   also used in gene therapy. The present sequence is R. toruloides PAL
CC   protein. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ   Sequence 716 AA;

  Query Match             89.8%;  Score 3021;  DB 5;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 4.5e-300;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| |||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              |||||||||||||||||||||||||:|||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| ||||||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:|||||||| |    |: ||||   |      ||  || | | |||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||     :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:||||||||||||||||||||||||||||||:|||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

RESULT 6
AAE20653
ID   AAE20653 standard; protein; 716 AA.
XX
AC   AAE20653;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Rhodotorula glutinis phenylalanine ammonia lyase (PAL).
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst.
XX
OS   Rhodotorula glutinis.
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   N-PSDB; AAD33075.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Example 6; Page 92-95; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL protein
XX
SQ   Sequence 716 AA;

  Query Match             89.8%;  Score 3021;  DB 5;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 4.5e-300;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| |||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173
```

```
Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||:||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||| |||||||||||||||| |:||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||| |       |: ||||    |          ||  ||  |  | ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              |||||||| ||||| |||       :||||||||||| || |||||| || || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:||||||||||||||||||||||||||||||:|||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

RESULT 7
AAE16389

```
ID   AAE16389 standard; protein; 716 AA.
XX
AC   AAE16389;
XX
DT   09-APR-2002  (first entry)
XX
DE   Rhodotorula glutinis wildtype phenylalanine ammonia-lyase (PAL).
XX
KW   Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
KW   para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW   Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW   aerospace application; chemical; red yeast.
XX
OS   Rhodotorula glutinis.
XX
PN   WO200111071-A2.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US021156.
XX
PR   06-AUG-1999;    99US-0147719P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Tang X,  Vannelli TM,  Qi WW,  Sariaslani S,  Gatenby AA;
XX
DR   WPI; 2002-121549/16.
DR   N-PSDB; AAD26923.
XX
PT   Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT   comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia
PT   -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT   activity.
XX
PS   Claim 8; Page 65-67; 75pp; English.
XX
CC   The invention relates to a method for producing para-hydroxycinnamic acid
CC   (PHCA). The method comprises converting cinnamate to PHCA by converting
CC   glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC   route or by generating a new biocatalyst possessing enhanced tyrosine
CC   ammonia-lyase (TAL) activity. The method is useful for the biological
CC   production of PHCA which is useful as a monomer for the production of
CC   Liquid Crystal Polymers (LCP), where the LCP is used in electronic
CC   connectors, telecommunications and aerospace applications. LCP is also
CC   useful in medical devices, as well as chemical, and food packing
CC   applications due to its resistance to sterilising radiation. The present
CC   sequence is red yeast, Rhodotorula glutinis (Rhodosporidium toruloides)
CC   wildtype PAL
XX
SQ   Sequence 716 AA;

  Query Match             89.8%;  Score 3021;  DB 5;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 4.5e-300;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||  :    |    ||          | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||    |||||| ||||| |||||| ||| |||||| ||||||||| ||   |||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||:| ||| |||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||:||||||::||||||||||||||||| :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||| |||||||||||||||| |:||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||| |       |: ||||    |          ||  ||  |  | ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              |||||||| ||||| |||       :||||||||||| || |||||| || || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::||:||||||||||||||||||||||||||||||:|||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710
```

```
Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

```
RESULT 8
ABG71948
ID   ABG71948 standard; protein; 716 AA.
XX
AC   ABG71948;
XX
DT   27-JAN-2003  (first entry)
XX
DE   R. glutinis wild-type Phenylalanine ammonia-lyase, PAL.
XX
KW   Enzyme; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase; PAH;
KW   para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; tyrosine;
KW   cinnamic acid; PHCA; liquid crystal polymer.
XX
OS   Rhodtorula glutinis.
XX
PN   WO200290523-A2.
XX
PD   14-NOV-2002.
XX
PF   03-MAY-2002; 2002WO-US018551.
XX
PR   04-MAY-2001; 2001US-0288701P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Qi WW,  Sariaslani FS,  Tang X;
XX
DR   WPI; 2003-058928/05.
DR   N-PSDB; ABS56530.
XX
PT   Novel recombinant host for producing para-hydroxycinnamic acid and
PT   tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT   gene encoding phenylalanine hydroxylase activity.
XX
PS   Claim 5; Page 48-50; 69pp; English.
XX
CC   The invention relates to a recombinant host comprising at least one gene
CC   encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
CC   encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
CC   (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC   providing the recombinant host; and (2) growing the recombinant organism
CC   in the presence of a fermentable carbon substrate and thus producing para
CC   -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC   providing a recombinant organism comprising at least one gene encoding a
CC   phenylalanine hydroxylase activity; and (2) growing the recombinant
CC   organism in the presence of a fermentable carbon substrate to produce
CC   tyrosine. The recombinant host cell is used for producing para-
CC   hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC   production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC   monomer for the production of liquid crystal polymer. The present
CC   sequence is R. glutinis wild-type Phenylalanine ammonia-lyase, PAL, an
CC   enzyme which catalyses the conversion of phenylalanine into cinnamic acid
XX
SQ   Sequence 716 AA;

  Query Match             89.8%;  Score 3021;  DB 6;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 4.5e-300;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||:  |  |||   :     |     ||             |  |||  ||||    ||:||||   ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              |  ||    ||||||| |||||  |||||| ||| |||||| |||||||||| ||   |||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||:|  |||  ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || |||  |||: ||||:||| ||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| |||||  ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:|||||||::|||||||||||||||||  :||||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| |||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||||||||||| |||||||||||||||| :|||||||||||||||||
Db        474 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||||  |      |: ||||    |         ||   ||  |   | ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               ||||||| | |||||| |||       :||||||||||  ||  ||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| |||||::||:||||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716
```

```
RESULT 9
ADE36054
ID   ADE36054 standard; protein; 716 AA.
XX
AC   ADE36054;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rhodosporidium glutinis PAL protein SEQ ID NO:2.
XX
KW   recombination; nucleic acid template; recombined polypeptide;
KW   altered property; combined gene.
XX
OS   Rhodosporidium.
XX
PN   WO2003072743-A2.
XX
PD   04-SEP-2003.
XX
PF   26-FEB-2003; 2003WO-US005708.
XX
PR   26-FEB-2002; 2002US-0360279P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Milano J,  Tang X;
XX
```

```
DR   WPI; 2003-721766/68.
DR   N-PSDB; ADE36053.
XX
PT   Recombination of nucleic acid templates, useful e.g. for producing
PT   enzymes with altered properties, is based on cyclic extension of unpaired
PT   primers.
XX
PS   Disclosure; SEQ ID NO 2; 481pp; English.
XX
CC   The present invention describes a method for the recombination of nucleic
CC   acid (NA) templates. The method comprises: (a) providing at least two
CC   double-stranded templates (T1, T2) having different 5' and 3' regions in
CC   their sense strands; (b) contacting the templates with at least one each
CC   of forward and reverse primers that anneal, respectively, only to the 3'-
CC   region of the antisense strand of T1 and to the 3'-region of the sense
CC   strand of T2; (c) extending the primers by no more than 1000 nucleotides
CC   (nt); (d) separating extended primers from their templates; (e)
CC   reannealing the extended primers, to either template; and (f) repeating
CC   cycles of steps (b)-(e) until at least one full-length extension product,
CC   i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
CC   method starting from at least one each of antisense and sense single-
CC   stranded templates; (2) generating a recombined polypeptide (I) with
CC   altered properties by expressing recombination products of the new
CC   processes and screening the polypeptides formed for properties different
CC   from those of polypeptides from either of the templates; and (3) (I)
CC   produced by method (2). The methods are useful for the recombination of
CC   nucleic acid templates of interest, and generation of a recombined
CC   polypeptide having altered properties. The method can be used to make
CC   combined genes that express proteins e.g. enzymes, cytokines, growth
CC   factors, viral proteins or microbial antigens with altered properties,
CC   e.g. stability, activity or specificity. The present sequence is used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 716 AA;

  Query Match             89.8%;  Score 3021;  DB 7;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 4.5e-300;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||  :   |   ||           | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   ||||||| ||||| |||||| ||| |||||| ||||||||| ||  |||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:||||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:|||||||::||||||||||||||||||| :|||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| |||||||||||||| |:|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:|||||||| |     |: ||||   |        || || | | ||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               ||||||| | ||||| |||      :|||||||||| || ||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| |||||::||:|||||||||||||||||||||||||||||||:|||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716


RESULT 10
AAE20668
ID   AAE20668 standard; protein; 716 AA.
XX
AC   AAE20668;
XX
DT   01-JUL-2002  (first entry)
XX
DE   R. glutinis phenylalanine ammonia lyase (PAL) mutant, EP18Km-6.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; mutant; variant; mutein.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key                Location/Qualifiers
FT   Misc-difference    540
FT                      /note= "Wild-type Ile substituted with Thr"
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   N-PSDB; AAD33100.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Disclosure; Page; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
```

```
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   mutant PAL. Note: This sequence is not shown in the specification,
CC   however it is constructed based on the PAL enzyme shown as SEQ.ID.NO:8
CC   (AAE20653) in the sequence listing
XX
SQ   Sequence 716 AA;

  Query Match             89.7%;  Score 3019;  DB 5;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.2e-300;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |  ||          | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   |||||| ||||| ||||| ||| |||||| ||||||||| ||  |||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||:| ||| ||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:||||||::||||||||||||||||  :|||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||| |||||||||||||||| |:||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:|||| |||||||| |     |: |||     |         ||  || | | ||||||||||
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| | ||||| |||        :||||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||::|:||||||||||||||||||||||||||||||:||||||||||| |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

RESULT 11
AAE20654
ID   AAE20654 standard; protein; 716 AA.
XX
AC   AAE20654;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant.
XX
KW   Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW   phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW   cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW   LCP; telecommunication; medical device; aerospace application; enzyme;
KW   biocatalyst; mutant; variant; mutein.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
PN   WO200210407-A1.
XX
PD   07-FEB-2002.
XX
PF   23-JAN-2001; 2001WO-US002099.
XX
PR   27-JUL-2000; 2000US-00627216.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gatenby AA,  Sariaslani FS,  Tang X,  Qi WW,  Vannelli T;
XX
DR   WPI; 2002-280635/32.
DR   N-PSDB; AAD33076.
XX
PT   Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT   polypeptide, or mutant TAL polypeptide, useful for producing
PT   parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT   hydroxylase.
XX
PS   Claim 2; Page 98-101; 139pp; English.
XX
CC   The invention relates to methods for biological production of para-
CC   hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC   biocatalyst for conversion of glucose to PHCA by incorporation of the
CC   wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC   Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC   wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC   to developing a new biocatalyst for conversion of glucose to PHCA by
CC   incorporation of the wildtype PAL which possesses enhanced tyrosine
CC   ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC   plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC   Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC   recombinant techniques. The recombinantly produced PHCA may be used as a
CC   monomer for production of liquid crystal polymers (LCP). LCP may be used
CC   in electronic connectors and telecommunication and aerospace
CC   applications. LCP resistance to sterilising radiation has also enabled
CC   these materials to be used in medical devices as well as chemical, and
CC   food packaging applications. The present sequence is Rhodotorula glutinis
CC   PAL mutant
XX
SQ   Sequence 716 AA;

  Query Match             89.7%;  Score 3019;  DB 5;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.2e-300;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              ||||||||: | |||  :   |  ||          | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   |||||| ||||| ||||| ||| |||||| ||||||||| ||  |||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
```

```
               ||||||||||||||||||||||||||||||:|  |||  ||||||||||||||||||||
Db       114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy       181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db       174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy       241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
             || |||  |||:  ||||:||| ||||||||||||||||||||||||||||||||||||
Db       234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy       301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
             |:|||:|||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db       294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy       361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
             |||||||||||||||||||||||:|||||||::|||||||||||||||| :||||||
Db       354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy       421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
             ||:|:| ||||||| || ||||||||||||||||||||||||||||||||||||||||
Db       414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy       481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
             |||||||||||||||||||||| ||||||||||||| |:||||||||||||||||
Db       474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy       541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
             |:|||| ||||||| |      |: ||||    |        ||  || | | ||||||||||
Db       534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy       601 EPRWHDAFSXATGTVVEXLSSXXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
              |||||||| | ||||| |||      :||||||||| || |||||| ||  || | |:
Db       594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy       661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
             ||||| ||||||::||:||||||||||||||||||||||||||:||||||||||| |
Db       651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy       721 LVKMLA 726
             |:||||
Db       711 LLKMLA 716

RESULT 12
AAE16390
ID   AAE16390 standard; protein; 716 AA.
XX
AC   AAE16390;
XX
DT   09-APR-2002  (first entry)
XX
DE   Rhodotorula glutinis mutant phenylalanine ammonia-lyase (PAL).
XX
KW   Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
KW   para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW   Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW   aerospace application; chemical; red yeast; mutant; mutein.
XX
OS   Rhodotorula glutinis.
OS   Synthetic.
XX
FH   Key                 Location/Qualifiers
FT   Misc-difference     540
FT                       /note= "Wild type Ile substituted with Thr"
XX
PN   WO200111071-A2.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US021156.
XX
PR   06-AUG-1999;   99US-0147719P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Tang X,  Vannelli TM,  Qi WW,  Sariaslani S,  Gatenby AA;
XX
DR   WPI; 2002-121549/16.
DR   N-PSDB; AAD26924.
XX
PT   Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT   comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia
PT   -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT   activity.
XX
PS   Claim 14; Page 68-70; 75pp; English.
XX
CC   The invention relates to a method for producing para-hydroxycinnamic acid
CC   (PHCA). The method comprises converting cinnamate to PHCA by converting
CC   glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC   route or by generating a new biocatalyst possessing enhanced tyrosine
CC   ammonia-lyase (TAL) activity. The method is useful for the biological
CC   production of PHCA which is useful as a monomer for the production of
CC   Liquid Crystal Polymers (LCP), where the LCP is used in electronic
CC   connectors, telecommunications and aerospace applications. LCP is also.
CC   useful in medical devices, as well as chemical, and food packing
CC   applications due to its resistance to sterilising radiation. The present
CC   sequence is red yeast, Rhodotorula glutinis (Rhodosporidium toruloides)
CC   mutant PAL derived by substituting Ile at position 540 with Thr. Mutant
CC   PAL enzyme posseses enhanced TAL activity
XX
SQ   Sequence 716 AA;

  Query Match             89.7%;  Score 3019;  DB 5;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.2e-300;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy         1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
             ||||||||: | |||  :   |   ||       | ||| ||||  ||:|||| ||
Db         1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy        61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
             | ||  |||||| ||||| ||||||  ||| |||||||||| || ||||||
Db        55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy       121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
             ||||||||||||||||||||||||||||||:|  |||  ||||||||||||||||||||
Db       114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy       181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db       174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy       241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
             || |||  |||:  ||||:||| ||||||||||||||||||||||||||||||||||||
Db       234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy       301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
             |:|||:|||||||||||||||||||||||||||||| ||| ||||| ||||||||||||
Db       294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy       361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
             |||||||||||||||||||||||:|||||||::|||||||||||||||| :||||||
Db       354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy       421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
             ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||
Db       414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy       481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
             |||||||||||||||||||||| ||||||||||||| |:||||||||||||||||
Db       474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy       541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
             |:|||| ||||||| |      |: ||||    |        ||  || | | ||||||||||
```

```
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726

Db        711 LLKMLA 716
```

```
RESULT 13
ABG71955
ID   ABG71955 standard; protein; 716 AA.
XX
AC   ABG71955;
XX
DT   27-JAN-2003  (first entry)
XX
DE   R. glutinis PAL/TAL naturally occurring variant.
XX
KW   Enzyme; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase; PAH;
KW   para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
KW   tyrosine; cinnamic acid; PHCA; liquid crystal polymer.
XX
OS   Rhodtorula glutinis.
XX
PN   WO200290523-A2.
XX
PD   14-NOV-2002.
XX
PF   03-MAY-2002; 2002WO-US018551.
XX
PR   04-MAY-2001; 2001US-0288701P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Qi WW,  Sariaslani FS,  Tang X;
XX
DR   WPI; 2003-058928/05.
DR   N-PSDB; ABS56543.
XX
PT   Novel recombinant host for producing para-hydroxycinnamic acid and
PT   tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT   gene encoding phenylalanine hydroxylase activity.
XX
PS   Disclosure; Page 67-69; 69pp; English.
XX
CC   The invention relates to a recombinant host comprising at least one gene
CC   encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
CC   encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
CC   (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC   providing the recombinant host; and (2) growing the recombinant organism
CC   in the presence of a fermentable carbon substrate and thus producing para
CC   -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC   providing a recombinant organism comprising at least one gene encoding a
CC   phenylalanine hydroxylase activity; and (2) growing the recombinant
CC   organism in the presence of a fermentable carbon substrate to produce
CC   tyrosine. The recombinant host cell is used for producing para-
CC   hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC   production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC   monomer for the production of liquid crystal polymer. The present
CC   sequence is a naturally occurring variant of the R. glutinis
CC   Phenylalanine ammonia-lyase, PAL, which has TAL activity (PAL/TAL)
XX
SQ   Sequence 716 AA;

  Query Match             89.7%;  Score 3019;  DB 6;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.2e-300;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240

Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300

Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360

Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420

Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Db        474 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726

Db        711 LLKMLA 716
```

```
RESULT 14
ADE36056
ID   ADE36056 standard; protein; 716 AA.
XX
AC   ADE36056;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Rhodosporidium glutinis EP18Km-6 mutant PAL protein SEQ ID NO:4.
XX
KW   recombination; nucleic acid template; recombined polypeptide;
KW   altered property; combined gene.
XX
OS   Rhodosporidium.
XX
PN   WO2003072743-A2.
XX
PD   04-SEP-2003.
```

```
XX
PF   26-FEB-2003; 2003WO-US005708.
XX
PR   26-FEB-2002; 2002US-0360279P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Milano J,  Tang X;
XX
DR   WPI; 2003-721766/68.
DR   N-PSDB; ADE36055.
XX
PT   Recombination of nucleic acid templates, useful e.g. for producing
PT   enzymes with altered properties, is based on cyclic extension of unpaired
PT   primers.
XX
PS   Disclosure; SEQ ID NO 4; 481pp; English.
XX
CC   The present invention describes a method for the recombination of nucleic
CC   acid (NA) templates. The method comprises: (a) providing at least two
CC   double-stranded templates (T1, T2) having different 5' and 3' regions in
CC   their sense strands; (b) contacting the templates with at least one each
CC   of forward and reverse primers that anneal, respectively, only to the 3'-
CC   region of the antisense strand of T1 and to the 3'-region of the sense
CC   strand of T2; (c) extending the primers by no more than 1000 nucleotides
CC   (nt); (d) separating extended primers from their templates; (e)
CC   reannealing the extended primers, to either template; and (f) repeating
CC   cycles of steps (b)-(e) until at least one full-length extension product,
CC   i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
CC   method starting from at least one each of antisense and sense single-
CC   stranded templates; (2) generating a recombined polypeptide (I) with
CC   altered properties by expressing recombination products of the new
CC   processes and screening the polypeptides formed for properties different
CC   from those of polypeptides from either of the templates; and (3) (I)
CC   produced by method (2). The methods are useful for the recombination of
CC   nucleic acid templates of interest, and generation of a recombined
CC   polypeptide having altered properties. The method can be used to make
CC   combined genes that express proteins e.g. enzymes, cytokines, growth
CC   factors, viral proteins or microbial antigens with altered properties,
CC   e.g. stability, activity or specificity. The present sequence is used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 716 AA;

  Query Match             89.7%;  Score 3019;  DB 7;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.2e-300;
  Matches  610;  Conservative   24;  Mismatches   82;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        474 AAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
Db        534 AIDLRATEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
Db        711 LLKMLA 716


RESULT 15
AAP81099
ID   AAP81099 standard; protein; 716 AA.
XX
AC   AAP81099;
XX
DT   12-DEC-1990  (first entry)
XX
DE   Sequence of Rhodosporidium toruloides IFO 559 cells L-Phenylalanine
DE   ammonia lyase (PAL).
XX
KW   Aspartame; artificial sweetener; enzyme.
XX
OS   Rhodosporidium toruloides.
XX
PN   EP260919-A.
XX
PD   23-MAR-1988.
XX
PF   15-SEP-1987;   87EP-00308119.
XX
PR   16-SEP-1986;   86JP-00215864.
XX
PA   (MITK ) MITSUI TOATSU CHEM INC.
XX
PI   Fukuhara N,  Yoshino S,  Yamamoto K,  Se T,  Sone S,  Nakajima Y;
PI   Suzuki M,  Makiguchi N;
XX
DR   WPI; 1988-079223/12.
DR   N-PSDB; AAN81399.
XX
PT   New L-phenylalanine ammonia lyase and encoding DNA sequence - useful for
PT   producing L-phenylalanine from cinnamic acid and ammonia.
XX
PS   Claim 1; Page 21-23; 37pp; English.
XX
CC   PAL is useful in prodn. of L-phenylalanine from cinnamic acid and an NH3
CC   donor (process claimed). L-Phe is an essential AA and starting material
CC   for sweetener aspartame. E.coli transformants will produce PAL without
CC   induction by expensive AAs, no surfactant is needed to increase cell wall
CC   permeability
XX
SQ   Sequence 716 AA;

  Query Match             89.7%;  Score 3018;  DB 1;  Length 716;
```

```
Best Local Similarity   83.9%;  Pred. No. 9.1e-300;
Matches  609;  Conservative   26;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||  :   |   ||           | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||   |||||| ||||| ||||| ||| |||||| ||||||||| ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              ||||||||||||||||||||||||||||||:| ||| |||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:|||||||||||||||||||||||||||| ||| ||||| ||||||||:|||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEQVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||||||||||||||:||||||::|||||||||||||||||||| :|||||
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
              ||:|| ||||||| || ||||||||||||||||||||||||||||||||||||||||||||
Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              |||||||||||||||||||||||| |||||||||||||||| :|||||||||||||||||
Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              |:||||:||||||||  |      |: ||||    |         ||  ||  |  | |||||||||||
Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               |||||||| ||||| |||       :|||||||||| || |||||| ||  || | |:
Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              |||| ||||||::||:||||||||||||||||||||||||||||||:|||||||||||  |
Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726
              |:||||
Db        711 LLKMLA 716

Search completed: September  9, 2004, 10:04:03
Job time : 92.3693 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:          September  9, 2004, 09:58:38 ; Search time 29.1203 Seconds
                                             (without alignments)
                                             2398.154 Million cell updates/sec

Title:           US-09-939-408A-21
Perfect score:   3365
Sequence:        1 MAPSLDSIATSXANGXXNGX..........RIYEAIKSGRINXVLVKMLA 726

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_78:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3021	89.8	716	2	A56628	phenylalanine ammo
2	2951	87.7	693	2	A29607	phenylalanine ammo
3	2922.5	86.8	713	2	S01999	phenylalanine ammo
4	969	28.8	725	2	G84787	phenylalanine ammo
5	965	28.7	725	2	S52990	phenylalanine ammo
6	958.5	28.5	701	2	S06475	phenylalanine ammo
7	957	28.4	754	2	T09777	phenylalanine ammo
8	955.5	28.4	718	2	S48726	phenylalanine ammo
9	954	28.4	725	2	S17444	phenylalanine ammo
10	946.5	28.1	716	2	S48725	phenylalanine ammo
11	946.5	28.1	716	2	S04463	phenylalanine ammo
12	946	28.1	713	2	S22991	phenylalanine ammo
13	943	28.0	715	2	S66343	phenylalanine ammo
14	939.5	27.9	717	2	T46172	phenylalanine ammo
15	938	27.9	712	2	T03663	phenylalanine ammo
16	932	27.7	723	2	S25303	phenylalanine ammo
17	931	27.7	721	2	A44133	phenylalanine ammo
18	929	27.6	667	2	T12749	phenylalanine ammo
19	929	27.6	712	2	T01858	phenylalanine ammo
20	928.5	27.6	712	2	S04127	phenylalanine ammo
21	927.5	27.6	708	2	T10909	phenylalanine ammo
22	927	27.5	708	2	T14295	phenylalanine ammo
23	927	27.5	710	2	S66313	phenylalanine ammo
24	925	27.5	705	2	JC5873	phenylalanine ammo
25	922	27.4	710	2	JC5872	phenylalanine ammo
26	918	27.3	590	2	S70916	phenylalanine ammo
27	916.5	27.2	720	2	S21174	phenylalanine ammo
28	910.5	27.1	715	2	JQ2265	phenylalanine ammo
29	910	27.0	695	2	S52992	phenylalanine ammo
30	910	27.0	710	2	S04128	phenylalanine ammo
31	906.5	26.9	710	2	S60042	phenylalanine ammo
32	898.5	26.7	549	2	T05970	phenylalanine ammo
33	893.5	26.6	707	2	S29029	phenylalanine ammo
34	892.5	26.5	619	2	S52632	phenylalanine ammo
35	891	26.5	700	2	T06545	phenylalanine ammo
36	874	26.0	716	2	S28185	phenylalanine ammo
37	841	25.0	582	2	S60043	phenylalanine ammo
38	664	19.7	506	2	S04129	phenylalanine ammo
39	659	19.6	505	2	A24727	phenylalanine ammo
40	612.5	18.2	497	2	T05968	probable phenylala
41	604.5	18.0	475	2	T05966	phenylalanine ammo
42	546.5	16.2	524	2	F75610	histidine ammonia-
43	543.5	16.2	514	2	AG3040	histidine ammonia-
44	543.5	16.2	514	2	E98245	hutH gene homolog
45	526.5	15.6	511	2	E82228	histidine ammonia-

ALIGNMENTS

```
RESULT 1
A56628
phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodosporidium toruloides) (strain CB
C;Species: Rhodosporidium toruloides
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C;Accession: A56628
R;Rasmussen, O.F.; Oerum, H.
DNA Seq. 1, 207-211, 1991
A;Title: Analysis of the gene for phenylalanine ammonia-lyase from Rhodosporidium torulo
A;Reference number: A56628; MUID:92127061; PMID:1773059
A;Accession: A56628
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-716 <RAS>
A;Cross-references: GB:X51513; NID:g3293; PIDN:CAA35886.1; PID:g3294
A;Note: sequence extracted from NCBI backbone (NCBIP:78381)
C;Genetics:
A;Gene: pal
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;212/Modified site: dehydroalanine (Ser) #status predicted

  Query Match             89.8%;  Score 3021;  DB 2;  Length 716;
  Best Local Similarity   84.0%;  Pred. No. 7.6e-230;
  Matches  610;  Conservative   25;  Mismatches   81;  Indels   10;  Gaps    3;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60
              |||||||: | |||   :    |   ||         | ||| ||||  ||:||||  ||
Db          1 MAPSLDSISHSFANGVASAKQAVNGAS------TNLAVAGSHLPTTQVTQVDIVEKMLAA 54

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120
              | ||  |||||| |||| |||||| ||| |||||| ||||||||||| || ||  ||||||
Db         55 P-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSMSVYGVT 113

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180
              |||||||||||||||||||||||||||||||:| ||| |||||||||||||||||||||||
Db        114 TGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRV 173

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db        174 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAISGHPDSKV 233

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              || ||| |||: ||||:||| ||||||||||||||||||||||||||||||||||||||||
Db        234 HVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 293

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |:|||:||||||||||||||||||||||||||||||| ||| ||||| |||||||||||||
Db        294 QSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKD 353

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
```

```
Db        354 DEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNF 413

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480

Db        414 QAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 473

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540

Db        474 AAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQ 533

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600

Db        534 AIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDL 593

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660

Db        594 VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESAISLTRQVRETFWSAAST 650

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720

Db        651 SSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNV 710

Qy        721 LVKMLA 726

Db        711 LLKMLA 716
```

RESULT 2
A29607
phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodosporidium toruloides) (strain IF
C;Species: Rhodosporidium toruloides
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 16-Jul-1999
C;Accession: A29607
R;Anson, J.G.; Gilbert, H.J.; Oram, J.D.; Minton, N.P.
Gene 58, 189-199, 1987
A;Title: Complete nucleotide sequence of the Rhodosporidium toruloides gene coding for p
A;Reference number: A29607; MUID:88112870; PMID:2828184
A;Accession: A29607
A;Molecule type: DNA
A;Residues: 1-693 <ANS>
A;Cross-references: GB:M18261; NID:g169745; PIDN:AAA33883.1; PID:g169746
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;188-190/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;189/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             87.7%;  Score 2951;  DB 2;  Length 693;
  Best Local Similarity   86.4%;  Pred. No. 2.4e-224;
  Matches  590;  Conservative   23;  Mismatches   66;  Indels    4;  Gaps    2;

Qy         44 PTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDK 103

Db         15 PTTQVTQVDIVEKMLAAP-TDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDK 73

Qy        104 SVEFLRXQLXNSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGL 163

Db         74 SVEFLRSQLSMSVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGL 133

Qy        164 ENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLS 223

Db        134 ENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLS 193

Qy        224 PLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSA 283

Db        194 PLSYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSA 253

Qy        284 SMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLL 343

Db        254 SMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLL 313

Qy        344 EGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTD 403

Db        314 EGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTD 373

Qy        404 NPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL 463

Db        374 NPLIDVENKTSHHGGNFQAAAVANTMEKTRLGLAQIGKLNFTQLTEMLNAGMNRGLPSCL 433

Qy        464 AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEA 523

Db        434 AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTES 493

Qy        524 NDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXK 583

Db        494 NDVLSLLLATHLYCVLQAIDLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEK 553

Qy        584 VXKXLXKRLEQTNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKA 643

Db        554 VNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNAWKVAAAESA 610

Qy        644 ISLTRXVRXXFWXAPSSSSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGS 703

Db        611 ISLTRQVRETFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGS 670

Qy        704 NVSRIYEAIKSGRINXVLVKMLA 726

Db        671 NVSKIYEAIKSGRINNVLLKMLA 693
```

RESULT 3
S01999
phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodotorula rubra)
C;Species: Rhodotorula rubra
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
C;Accession: S01999; S01990
R;Filpula, D.; Vaslet, C.A.; Levy, A.; Sykes, A.; Strausberg, R.L.
Nucleic Acids Res. 16, 11381, 1988
A;Title: Nucleotide sequence of gene for phenylalanine ammonia-lyase from Rhodotorula
A;Reference number: S01999; MUID:89083582; PMID:3205749
A;Accession: S01999
A;Molecule type: DNA
A;Residues: 1-713 <FIL>
A;Cross-references: EMBL:X13094; NID:g3284; PIDN:CAA31486.1; PID:g295942
R;Vaslet, C.A.; Strausberg, R.L.; Sykes, A.; Levy, A.; Filpula, D.
Nucleic Acids Res. 16, 11382, 1988
A;Title: cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes reveal ger
A;Reference number: S01990; MUID:89083583; PMID:3205750
A;Accession: S01990
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-674 <VAS>
A;Cross-references: EMBL:X13095
C;Genetics:
A;Introns: 110/2; 139/2; 372/3; 428/2; 482/3
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;217-219/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;218/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             86.8%;  Score 2922.5;  DB 2;  Length 713;
  Best Local Similarity   82.0%;  Pred. No. 4.4e-222;
  Matches  595;  Conservative   33;  Mismatches   85;  Indels   13;  Gaps    5;

Qy          1 MAPSLDSIATSXANGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLAD 60

Db          1 MAPSVDSIATSVANSLSNGLHAAAAAN-GGDVHKKTAGAGSLLPTTETTQLDIVERILAD 59

Qy         61 PXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSVYGVT 120

Db         60 AGATDQIKLDGYTLTLGDVVGAARRGRSVKVADSPHIREKIDASVEFLRTQLDNSVYGVT 119

Qy        121 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRV 180

Db        120 TGFGGSADTRTEDAISLQKALLEHQLCGVLPTSMDGFALGRGLENSLPLEVVRGAMTIRV 179
```

```
Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
              ||||||||||:||||||||||||||||||||||||||||||||||||||:||||||||||
Db        180 NSLTRGHSAVRIVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAASITGHPDSKV 239

Qy        241 HVXHEGXEKIMXAREAIALFGLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLS 300
              ||  :|  |||:||||| ||:|||||||||||||||||||||||||||| |||:||||:
Db        240 HV--DG--KIMSAQEAIALKGLQPVVLGPKEGLGLVNGTAVSASMATLALTDAHVLSLLA 295

Qy        301 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKD 360
              |||||||||||||||||||||||||||||||||||||||||||||||| :||||| |||||
Db        296 QALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSKYAVHHETEVKVKD 355

Qy        361 DEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNF 420
              ||||||||||||| |||||||||||||||||||||||||||||||||||:||| ||||| |
Db        356 DEGILRQDRYPLRCSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDLENKMTHHGGAF 415

Qy        421 QASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIA 480
               ||:| ||||||||:||:|:|::|||||||||||||| |||||||||||||||||||||
Db        416 MASSVGNTMEKTRLAVALMGKVSFTQLTEMLNAGMNRALPSCLAAEDPSLSYHCKGLDIA 475

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
              ||||||||||||||:|||||||||||||:|||||||||||||||||||||||||||||||
Db        476 AAAYTSELGHLANPVSTHVQPAEMGNQAINSLALISARRTAEANDVLSLLLATHLYCVLQ 535

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRLEQTNSYDL 600
              ||||||||| |    |    |||||   |:  ||    |  :  |||:| ||||
Db        536 AVDLRAMEFEHTKAFEPMVTELLKQHFGALAT-----AEVEDKVRKSIYKRLQQNNSYDL 590

Qy        601 EPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSS 660
               ||||  ||  ||| |||         ||||::|||||| |||||:|||  || ||||
Db        591 EQRWHDTFSVATGAVVEALAG---QEVSLASLNAWKVACAEKAIALTRSVRDSFWAAPSS 647

Qy        661 SSPALXYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXV 720
              ||||| ||||||||||||||||:|||||||||:||||||||:|||||||||||   |
Db        648 SSPALKYLSPRTRVLYSFVREEVGVKARRGDVYLGKQEVTIGTNVSRIYEAIKSGCIAPV 707

Qy        721 LVKMLA 726
              ||||:|
Db        708 LVKMMA 713
```

RESULT 4
G84787
phenylalanine ammonia lyase (PAL1) [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: G84787
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-725 <STO>
A;Cross-references: GB:AE002093; NID:g6598547; PIDN:AAD18156.2; GSPDB:GN00139
C;Genetics:
A;Gene: At2g37040
A;Map position: 2
C;Superfamily: histidine ammonia-lyase

```
  Query Match             28.8%;  Score 969;  DB 2;  Length 725;
  Best Local Similarity   35.5%;  Pred. No. 3.3e-68;
  Matches  258;  Conservative  120;  Mismatches  257;  Indels   92;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRX 99
              |:      : || |:  :|:           | | :||:|  |    :  .|  |:|  |:  |
Db         39 GAAAEQMKGSHLDEVKRMVAE-FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARA 97

Qy        100 KIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSF 157
               :: | :::   :       ||||||| ::  ||:: ::||| |:       |:  :: ::
Db         98 GVNASSDWVMESMNKGTDSYGVTTGFGATSHRRTKNGVALQKELIRFLNAGIFGSTKET- 156

Qy        158 XLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTIS 217
                 ::||    | || :|:|:| :| |  :|  :|||:|:|||:  |||  :||||||:
Db        157 ------SHTLPHSATRAAMLVRINTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTIT 210

Qy        218 ASGDLSPLSYIAAAITGHPDSKVHVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGL 275
              ||||| |||||  :|| |:||          |   :  :  | ||  | |:        | |||||  |
Db        211 ASGDLVPLSYIAGLLTGRPNSKA-TGPNG--EALTAEEAFKLAGISSGFFDLQPKEGLAL 267

Qy        276 VNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEV 335
              ||||| : ||:: | : ::||:|:: |:|:  | | |       |     |      : || |||
Db        268 VNGTAVGSGMASMVLFETNVLSVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEA 326

Qy        336 ARNIRTLLEGSXF----AVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHA 391
              |  :  :|:|| :          || :  |         :||||  |||||||| :   :  :|
Db        327 AAIMEHILDGSSYMKLAQKLHEMDPLQKP-----KQDRYALRTSPQWLGPQIE--VIRYA 379

Qy        392 VLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450
               |:|       |  ||||||     ||||| : :  :|: |||:| ||||  |  | :|:
Db        380 TKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL 439

Qy        451 LNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509
               :|    | ||||  | |:  :|||  |   || :||  |:|  |||  :||||||:|||  ||  || |
Db        440 VNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDV 499

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569
              ||| |||:|:|:|| |:| |:  | |  :|||||| :|    ::
Db        500 NSLGLISSRKTSEAVDILKLMSTTFLVAICQAVDLRHLEENLRQ--------------- 543

Qy        570 XXTXXXXXXELXXKV--------------XKXLXKRL--EQTNSYDLEPRWHDAFSXAT 612
                |          : |             | | |  :  | | : | :|          ||
Db        544 --TVKNTVSQVAKKVLTTGVNGELHPSRFCEKDLLKVVDREQVYTYADDP------CSAT 595

Qy        613 GTVVEXLSSXXXXXVSLAAVNA----------WKVAXAEKAIS--LTRXVRXXFWXAPS 659
               ::: |   |  | :|          |:  |: :  | |       :
Db        596 YPLIQKLRQ---VIVDHALINGESEKNAVTSIFHKIGAFEEELKAVLPKEVEAARAAYDN 652

Qy        660 SSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRIN 718
               :|   :  |:  || ||||||    ::  |:: :  |    ::: ||  |:|
Db        653 GTSAIPNRIKECRSYPLYRFVREELGT-----ELLTGEKVTSPGEEFDKVFTAICEGKII 707

Qy        719 XVLVKML 725
              ::: |
Db        708 DPMMECL 714
```

RESULT 5
S52990
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C;Accession: S52990; PQ0140
R;Wanner, L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
Plant Mol. Biol. 27, 327-338, 1995
A;Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
A;Reference number: S52990; MUID:95195160; PMID:7888622
A;Accession: S52990
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-725 <WAN>
A;Cross-references: EMBL:L33677; NID:g497418; PIDN:AAC18870.1; PID:g497419
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
R;Ohl, S.; Hedrick, S.A.; Chory, J.; Lamb, C.J.
Plant Cell 2, 837-848, 1990
A;Title: Functional properties of a phenylalanine ammonia-lyase promoter from Arabidopsis
A;Reference number: PQ0140; MUID:93005677; PMID:2152131
A;Accession: PQ0140
A;Molecule type: DNA
A;Residues: 1-240 <OHL>
A;Cross-references: GB:X62747; NID:g16408; PIDN:CAA44609.1; PID:g16409
A;Experimental source: Ecotype Columbia
C;Comment: Phenylalanine ammonia-lyase catalyzes the deamination of L-phenylalanine to t

```
C;Genetics:
A;Gene: PAL1
A;Introns: 143/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;212/Modified site: dehydroalanine (Ser) #status predicted

  Query Match             28.7%;  Score 965;  DB 2;  Length 725;
  Best Local Similarity   35.5%;  Pred. No. 6.9e-68;
  Matches  258;  Conservative  119;  Mismatches  258;  Indels   92;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRX 99
Db         39 GAAAEQMKGSHLDEVKRMVAE-FRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARA 97

Qy        100 KIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSF 157
Db         98 GVNASSDWVMESMNKGTDSYGVTTGFGATSHRRTKNGVALQKELIRFLNAGIFGSTKET- 156

Qy        158 XLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTIS 217
Db        157 ------SHTLPHSATRAAMLVRINTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTIT 210

Qy        218 ASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGL 275
Db        211 ASGDLVPLSYIAGLLTGRPNSKA-TGPNG--EALTAEEAFKLAGISSGFFDLQPKEGLAL 267

Qy        276 VNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEV 335
Db        268 VNGTAVGSGMASMVLFETNVLSVLAEILSAVFAEVMSGKP-EFTDHLTHRLKHHPGQIEA 326

Qy        336 ARNIRTLLEGSXF----AVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHA 391
Db        327 AAVMEHILDGSSYMKLAQKLHEMDPLQKP-----KQDRYALRTSPQWLGPQIE--VIRYA 379

Qy        392 VLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450
Db        380 TKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIRAIGKLMFAQFSEL 439

Qy        451 LNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509
Db        440 VNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDV 499

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569
Db        500 NSLGLISSRKTSEAVDILKLMSTTFLVAICQAVDLRHLEENLRQ---------------- 543

Qy        570 XXTXXXXXXELXXKV--------------XKXLXKRL--EQTNSYDLEPRWHDAFSXAT 612
Db        544 --TVKNTVSQVAKKVLTTGVNGELHPSRFCEKDLLKVVDREQVYTYADDP------CSAT 595

Qy        613 GTVVEXLSSXXXXXVSLAAVNA-----------WKVAXAEKAIS--LTRXVRXXFWXAPS 659
Db        596 YPLIQKLRQ---VIVDHALVNGESEKNAVTSIFHKIGAFEEELKAVLPKEVEAARAAYDN 652

Qy        660 SSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRIN 718
Db        653 GTSAIPNRIKECRSYPLYRFVREELGT-----ELLTGEKVTSPGEEFDKVFTAICEGKII 707

Qy        719 XVLVKML 725
Db        708 DPMMECL 714

RESULT 6
S06475
phenylalanine ammonia-lyase (EC 4.3.1.5) - rice
C;Species: Oryza sativa (rice)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 21-May-1999
C;Accession: S06475; A33246; JQ0156
R;Minami, E.; Ozeki, Y.; Matsuoka, M.; Koizuka, N.; Tanaka, Y.
Eur. J. Biochem. 185, 19-25, 1989
A;Title: Structure and some characterization of the gene for phenylalanine ammonia-lya
A;Reference number: JQ0156; MUID:90032676; PMID:2806257
A;Accession: S06475
A;Molecule type: DNA
A;Residues: 1-701 <MIN>
A;Cross-references: EMBL:X16099
A;Note: the authors translated the codon GTC for residue 66 as Leu, CAG for residue 35
A;Accession: A33246
A;Molecule type: mRNA
A;Residues: 1-86,'N',88-302,'E',304-701 <MIN2>
C;Genetics:
A;Introns: 122/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;189-191/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;190/Modified site: dehydroalanine (Ser) #status predicted

  Query Match             28.5%;  Score 958.5;  DB 2;  Length 701;
  Best Local Similarity   36.0%;  Pred. No. 2.1e-67;
  Matches  255;  Conservative  116;  Mismatches  266;  Indels   71;  Gaps   21;

Qy         40 GSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYSLTLGDV--VGAARKGRXVRVXDSDEI 97
Db         16 GAAAAEMAGSHLDEVKRMVAQ-FREPLVKIQGATLRVGQVAAVAQAKDAARVAVELDEEA 74

Qy         98 RXKIDKSVEFLRXQLXN--SVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXD 155
Db         75 RPRVKASSEWILTCIAHGGDIYGVTTGFGGTSHRRTKDGPALQVELLRHLNAGIFGTGSD 134

Qy        156 SFXLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGT 215
Db        135 G--------HTLPSETVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVTPCLPLRGT 186

Qy        216 ISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGL 273
Db        187 ITASGDLVPLSYIAGLITGRPNAQA-ISPDGRK--VDAAEAFKLAGIEGGFFTLNPKEGL 243

Qy        274 GLVNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQI 333
Db        244 AIVNGTSVGSALAATVMFDANILAVLSEVLSAVFCEVMNGKPEYTDHLTHKL-KHHPGSI 302

Qy        334 EVARNIRTLLEGSXFAVHHEEEVKVKDDEGIL--RQDRYPLRTSPQWLGPLVSDMIHAHA 391
Db        303 DAAAIMEHILAGSSFMSHAK---KVNEMDPLLKPKQDRYALRTSPQWLGPQI-QVIRA-A 357

Qy        392 VLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEM 450
Db        358 TKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSEL 417

Qy        451 LNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAV 509
Db        418 VNEFYNNGLTSNLAGSRNPSLDYGFKGTEIAMASYSSELQYLANPITNHVQSAEQHNQDV 477

Qy        510 NSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGX 569
Db        478 NSLGLVSARKTLEAVDILKLMTSTYIVALCQAVDLRHLEENIKSSVKNCVTQVAKKVLTM 537

Qy        570 XXTXXXXXXELXXKVXKXLXKRLEQTNSYDLEP--RWHDAFSXATGTVVE-XLSSXXXXX 626
Db        538 NPTGDLSSARFSEKNLLTAIDR-EAVFSYADDPCSANYPLMQKLRAVLVEHALTSGDRRA 596

Qy        627 VSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSSSSP----------------ALXYLSP 670
Db        597 RGLRVLQDHQVRGG----------------APLCAAPGDRGRPRRRQRTAPVANRIVES 640

Qy        671 RTRVLYSFVREELGVKARRGDVFL-GKQEVTIGSNVSRIYEAIKSGRI 717
Db        641 RSFPLYRFVREELGC------VFLTGEKLKSPGEECNKVFLGISQGKL 682

RESULT 7
T09777
phenylalanine ammonia-lyase (EC 4.3.1.5) - loblolly pine
```

```
C;Species: Pinus taeda (loblolly pine)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 31-Jan-2000
C;Accession: T09777
R;Zhang, X.H.; Chiang, V.L.
submitted to the EMBL Data Library, October 1995
A;Description: Nucleotide sequence of loblolly pine phenylalanine ammonia-lyase gene, lp
A;Reference number: Z16847
A;Accession: T09777
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-754 <ZHA>
A;Cross-references: EMBL:U39792; NID:g1143311; PID:g1143312
C;Genetics:
A;Gene: PAL
C;Function:
A;Description: catalyzes the deamination of L-phenylalanine to trans-cinnamic acid
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;206-208/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;207/Modified site: dehydroalanine (Ser) #status predicted

  Query Match            28.4%;  Score 957;  DB 2;  Length 754;
  Best Local Similarity  36.9%;  Pred. No. 3.1e-67;
  Matches  252;  Conservative  109;  Mismatches  258;  Indels  64;  Gaps  18;

Qy         69 LDGYSLTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLX--NSVYGVTTGFGGS 126
Db         64 IEGKSLTISDVAAVARRSQVKVKLDAAAAKSRVEESSNWVLTQMTKGTDTYGVTTGFGAT 123

Qy        127 ADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTRG 186
Db        124 SHRRTNQGAELQKELIRFLNAGV---------LGKCPENVLSEDTTRAAMLVRTNTLLQG 174

Qy        187 HSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEG 246
Db        175 YSGVRWDILETVEKLLNAWLTPKLPLRGTITASGDLVPLSYIAGLLTGRPNSRVR-SRDG 233

Qy        247 XEKIMXAREAIALFGLE-PVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTA 305
Db        234 IE--MSGAEALKKVGLEKPFELQPKEGLAIVNGTSVGAALASIVCFDANVLALLSEVISA 291

Qy        306 LTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH----HEEVKVKDD 361
Db        292 MFCEVMNGKPEFTDPLTHKL-KHHPGQMEAAAIMEYVLDGSSYMKHAAKLHEMNPLQKP- 349

Qy        362 EGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQ 421
Db        350 ----KQDRYGLRTSPQWLGPQVEIIRSATHMIEREI-NSVNDNPVIDVARDKALHGGNFQ 404

Qy        422 ASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIA 480
Db        405 GTPIGVSMDNLRLSISAIGKLMFAQFSELVNDYYNGGLPSNLSGGPNPSLDYGLKGAEIA 464

Qy        481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQ 540
Db        465 MASYTSELLYLANPVTSHVQSAEQHNQDVNSLGLVSARKSAEAIDILKLMLSTYLTALCQ 524

Qy        541 AVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXK-VXKXLXKRLEQTN--- 596
Db        525 AVDLRHLE----ENMLATVKQIVSQVAKKTLSTGLNGELLPGRFCEKDLLQVVDNEHVFS 580

Qy        597 --------SYDLEPRWHD-----AFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXAEKA 643
Db        581 YIDDPCNASYPLTQKLRNILVEHAFKNAEGEKDPNTSIFNKIPVFEAELK----AQLEPQ 636

Qy        644 ISLTRXVRXXFWXAPSSSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIG 702
Db        637 VSLARE------SYDKGTSPLPDRIQECRSYPLYEFVRNQLGTK-----LLSGTRTISPG 685

Qy        703 SNVSRIYEAIKSGRINXVLVKML 725
Db        686 EVIEVVYDAISEDKVIVPLFKCL 708
```

```
RESULT 8
S48726
phenylalanine ammonia-lyase (EC 4.3.1.5) 3 - parsley
C;Species: Petroselinum crispum (parsley)
C;Date: 18-Oct-1995 #sequence_revision 12-Apr-1996 #text_change 16-Jul-1999
C;Accession: S48726; S56035
R;Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
Eur. J. Biochem. 225, 491-499, 1994
A;Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase iso
A;Reference number: S48725; MUID:95010141; PMID:7925471
A;Accession: S48726
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-718 <APP1>
A;Cross-references: EMBL:X81159
R;Appert, C.
submitted to the EMBL Data Library, August 1994
A;Reference number: S56035
A;Accession: S56035
A;Molecule type: mRNA
A;Residues: 1-58,'E',60-718 <APP2>
A;Cross-references: EMBL:X81159; NID:g535007; PIDN:CAA57057.1; PID:g535008
C;Genetics:
A;Gene: PAL3
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;204-206/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;205/Modified site: dehydroalanine (Ser) #status predicted

  Query Match            28.4%;  Score 955.5;  DB 2;  Length 718;
  Best Local Similarity  35.2%;  Pred. No. 3.8e-67;
  Matches  264;  Conservative  114;  Mismatches  286;  Indels  87;  Gaps  21;

Qy         14 NGXXNGXHAAXXASXXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYS 73
Db          5 NGTTNG-HANGNGLDLCMKKEDPLNWGVAAEALTGSHLDEVKRMVAE-YRKPVVKLGGET 62

Qy         74 LTLGDVVG-AARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFGGSADTR 130
Db         63 LTISQVAAISARDDSGVKVELSEEARAGVKASSDWVMDSMNKGTDSYGVTTGFGATSHRR 122

Qy        131 TEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLE---NSLPLEVVRGAMTIRVNSLTRGH 187
Db        123 TKQGGALQKELIRFLNAGI---------FGSGAEAGNNTLPHSATRAAMLVRINTLLQGY 173

Qy        188 SAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEGX 247
Db        174 SGIRFEILEAITKFLNHNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPTGV 232

Qy        248 EKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTA 305
Db        233 --TLSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEVMSA 290

Qy        306 LTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKVKDDEGIL 365
Db        291 IFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSAY-VKAAQKLHEMDPLQKP 348

Qy        366 RQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGGNFQASAV 425
Db        349 KQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGSPI 407

Qy        426 XNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCLA-AEDPSLSYHCKGLDIAAAAY 484
Db        408 GVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASY 467

Qy        485 TSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQAVDL 544
Db        468 CSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAIDL 527

Qy        545 RAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKV--------------XKXLX 589
Db        528 RHLEENLK-----------------STVKNTVSQVAKRVLTMGVNGELHPSRFCEKDLL 569
```

```
Qy        590 KRLEQ----------TNSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVA 638
Db        570 RVVDREYIFAYIDDPCSATYPLMQKLRE-------TLVEHALNNGDKERNLSTSIFQKIA 622

Qy        639 XAEKAIS--LTRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSFVREELGVKARRGDVFL 694
Db        623 AFEDELKALLPKEVETA-RAALESGNPAIPNRIKECRSYPLYKFVREELGT-----EYLT 676

Qy        695 GKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
Db        677 GEKVRSPGEEFEKVFTAMSKGEIIDPLLECL 707
```

RESULT 9
S17444
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - alfalfa
C;Species: Medicago sativa (alfalfa)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S17444
R;Gowri, G.; Paiva, N.L.; Dixon, R.A.
Plant Mol. Biol. 17, 415-429, 1991
A;Title: Stress responses in alfalfa (Medicago sativa L.). 12. Sequence analysis of phen
nd developing plants.
A;Reference number: S17444; MUID:91355933; PMID:1715786
A;Accession: S17444
A;Molecule type: mRNA
A;Residues: 1-725 <GOW>
A;Cross-references: EMBL:X58180; NID:g19649; PIDN:CAA41169.1; PID:g19650
A;Experimental source: strain cv. Apollo
C;Genetics:
A;Gene: PAL1
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;212/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.4%;  Score 954;  DB 2;  Length 725;
  Best Local Similarity   36.2%;  Pred. No. 5.1e-67;
  Matches  262;  Conservative  106;  Mismatches  252;  Indels  104;  Gaps   21;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARKGRXVRVXDSDEIRXXIDKSVEFL 108
Db         48 SHLDEVKRMVAE-YRKPVVRLGGETLTISQVAAIAAHDHGVQVDLSESARDGVKASSEWV 106

Qy        109 RXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENS 166
Db        107 MESMNKGTDSYGVTTGFGATSHSRTKQGGALQKELIRFLNAGIFGNGTES-------NHT 159

Qy        167 LPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLS 226
Db        160 LPKTATRAAMLVRINTLLQGYSGIDFEILEAITKPLNKTVTPCLPLRGTITASGDLVPLS 219

Qy        227 YIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGL--EPVVLGPKEGLGLVNGTAVSAS 284
Db        220 YIAGLLTGRPNSK---AHGPSGEVLNAKEAFNLAGINAEFFELQPKEGLALVNGTAVGSG 276

Qy        285 MATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLE 344
Db        277 LASIVLFEANILAVLSEVLSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILD 335

Qy        345 GSXFAVH----HEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAG-Q 399
Db        336 GSSYVKAAKKLHEIDPLQKP-----KQDRYALRTSPQWLGPLVE--VIRFSTKSIEREIN 388

Qy        400 STTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGL 459
Db        389 SVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNTRLALASIGKLMFAQFSELVNDFYNNGL 448

Qy        460 PSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISAR 518
Db        449 PSNLSASRNPSLDYGFKGAEIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISAR 508

Qy        519 RTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXX 578
Db        509 KTNEAIEILQLMSSTFLIALCQAIDLRHLEENLKN------------------------- 543

Qy        579 ELXXKVXKXLXKRLEQTNSYDLEP------------RWH-----DAFSXATGTVVEXLS 620
Db        544 SVKNTVSQVAKKTLTMGVNGELHPSRFCEKDLLKVVDREHVFAYIDDPCSATYPLSQKLR 603

Qy        621 SXXXXXVSLAAVNA----------WKVAXAEKAIS--LTRXVRXXFWXAPSSSSPALXY 667
Db        604 Q---VLVDHALVNGESEKNFNTSIFQKIATFEEELKTLLPKEVESA-RTAYESGNPTI-- 657

Qy        668 LSP------RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVL 721
Db        658 --PNKINGCRSYPLYKFVREELGT-----GLLTGENVISPGEECDKLFSAMCQGKIIDPL 710

Qy        722 VKML 725
Db        711 LECL 714
```

RESULT 10
S48725
phenylalanine ammonia-lyase (EC 4.3.1.5) 2 - parsley
C;Species: Petroselinum crispum (parsley)
C;Date: 28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48725
R;Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
Eur. J. Biochem. 225, 491-499, 1994
A;Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase is
A;Reference number: S48725; MUID:95010141; PMID:7925471
A;Accession: S48725
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-716 <APP>
A;Cross-references: EMBL:X81158; NID:g534892; PIDN:CAA57056.1; PID:g534893
C;Genetics:
A;Gene: PAL2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;202-204/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;203/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.1%;  Score 946.5;  DB 2;  Length 716;
  Best Local Similarity   35.5%;  Pred. No. 1.9e-66;
  Matches  254;  Conservative  110;  Mismatches  264;  Indels   87;  Gaps   19;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVG-AARKGRXVRVXDSDEIRXXIDKSVEF 107
Db         40 SHLDEVKKMVAE-YRKPVVKLGGETLTISQVAAISARDSGVTVELSEAARAGVKASSDW 98

Qy        108 LRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLEN 165
Db         99 VMDSMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI---------FGNGSDN 149

Qy        166 SLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL 225
Db        150 TLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPL 209

Qy        226 SYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSA 283
Db        210 SYIAGLLTGRPNSKA-VGPTGV--ILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGS 266

Qy        284 SMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLL 343
Db        267 GMASMVLFEANILAVLAEVMSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHIL 325

Qy        344 EGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTD 403
Db        326 DGSAY-VKAAQKLHEMDPLQKPKQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINSVND 383

Qy        404 NPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL 463
```

```
Db        384 NPLIDVSRNKAIHGGNFQGTPIGMSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNL 443

Qy        464 A-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAE 522

Db        444 SGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSE 503

Qy        523 ANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHF-----GXXXTXXXXX 577

Db        504 AVEILKLMSTTFLVGLCQAIDLRHLEENLKSTVKNTVSSVAKRVLTMGVNGELHPSRFCE 563

Qy        578 XELXXKVXKX--------------LXKRLEQT------NSYDLEPRWHDAFSXATGTVV 616

Db        564 KDLLRFVDREYIFAYIDDPCSATYPLMQKLRQTLVEHALKNGDNERNMNTSIFQKIATFE 623

Qy        617 EXLSSXXXXXVSLAAVNAWKVAXAEKAISLTRXVRXXFWXAPSSSSPALXYLSP------ 670

Db        624 DELK----------------ALLPKEVESAR-------AALESGNPAI----PNRIEEC 655

Qy        671 RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725

Db        656 RSYPLYKFVRKELGI-----EYLTGEKVTSPGEEFDKVFIAMSKGEIIDPLLECL 705
```

RESULT 11
S04463
phenylalanine ammonia-lyase (EC 4.3.1.5) - parsley
C;Species: Petroselinum crispum (parsley)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-May-1999
C;Accession: S04463
R;Lois, R.; Dietrich, A.; Hahlbrock, K.; Schulz, W.
EMBO J. 8, 1641-1648, 1989
A;Title: A phenylalanine ammonia-lyase gene from parsley: structure, regulation and iden
A;Reference number: S04463; MUID:89356635; PMID:2767049
A;Accession: S04463
A;Molecule type: DNA
A;Residues: 1-716 <LOI>
A;Cross-references: EMBL:X15473
C;Genetics:
A;Introns: 136/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;202-204/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;203/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.1%;  Score 946.5;  DB 2;  Length 716;
  Best Local Similarity   36.7%;  Pred. No. 1.9e-66;
  Matches  254;  Conservative  110;  Mismatches  286;  Indels   43;  Gaps   18;

Qy         49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVG-AARKGRXVRVXDSDEIRXKIDKSVEF 107

Db         40 SHLDEVKKMVAE-YRKPVVKLGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDW 98

Qy        108 LRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLEN 165

Db         99 VMDSMNKGTDSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGI---------FGNGSDN 149

Qy        166 SLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPL 225

Db        150 TLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPL 209

Qy        226 SYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSA 283

Db        210 SYIAGLLTGRPNSKA-VGPTGV--ILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGS 266

Qy        284 SMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLL 343

Db        267 GMASMVLFEANILAVLAEVMSAIFAEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHIL 325

Qy        344 EGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTD 403

Db        326 DGSAY-VKAAQKLHEMDPLQKPKQDRYALRTSPQWLGPQI-EVIRSSTKMIEREINSVND 383

Qy        404 NPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL 463

Db        384 NPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDFYNNGLPSNL 443

Qy        464 A-AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAE 522

Db        444 SGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSE 503

Qy        523 ANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXX 582

Db        504 AVEILKLMSTTFLVGLCQAIDLRHLEENLKSTVKNTVSSVAKRVLTMGVNGELHPSRFCE 563

Qy        583 KVXKXLXKRLEQTNSYDLEP--RWHDAFSXATGTVVEXLSSXXXXXVSLAAVNAWKVAXA 640

Db        564 KDLLRVVDR-EYIFAYIDDPCSATYPLMQKLRQTLVEHALKNGDNERNLSTSIFQKIATF 622

Qy        641 EKAIS--LTRXVRXXFWXAPSSSSPALXYLSP------RTRVLYSFVREELGVKARRGDV 692

Db        623 EDELKALLPKEVESA-RAALESGNPAI----PNRIEECRSYPLYKFVRKELGT-----EY 672

Qy        693 FLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725

Db        673 LTGEKVTSPGEEFEKVFIAMSKGEIIDPLLECL 705
```

RESULT 12
S22991
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - soybean
C;Species: Glycine max (soybean)
C;Date: 06-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S22991
R;Frank, R.L.; Vodkin, L.O.
DNA Seq. 1, 335-346, 1991
A;Title: Sequence and structure of a phenylalanine ammonia-lyase gene from Glycine max.
A;Reference number: S22991; MUID:92190550; PMID:1799682
A;Accession: S22991
A;Molecule type: DNA
A;Residues: 1-713 <FRA>
A;Cross-references: EMBL:X52953; NID:g18376; PIDN:CAA37129.1; PID:g18377
C;Genetics:
A;Gene: PAL1
A;Introns: 131/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;199-201/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;200/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.1%;  Score 946;  DB 2;  Length 713;
  Best Local Similarity   36.6%;  Pred. No. 2.1e-66;
  Matches  271;  Conservative  105;  Mismatches  293;  Indels   72;  Gaps   23;

Qy         14 NGXXNGXHAAXXASXXXXXXXXXAXAGSXLPTTXXTQLDIVEXXLADPXTDDXXELDGYS 73

Db          5 NGHQNGSFCLSTAKGNNDPLNWGAAA----EAMKGSHLDEVKRMVAE-YRKPVVRLGGET 59

Qy         74 LTLGDVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRT 131

Db         60 LTIAQVAAVAGHDHGVAVELSESAREGVKASSEWVMNSMNNGTDSYGVTTGFGATSHRRT 119

Qy        132 EDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENS--LPLEVVRGAMTIRVNSLTRGHSA 189

Db        120 KQGGALQKELIRFLNAGI---------FGNGTESSHTLPHTATRAAMLVRINTLLQGYSG 170

Qy        190 VRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEK 249

Db        171 IRFEILEAITKLLNNNVTPCLDLRGTITASGDLVPLSYIAGLLTGRPNSKA-VGPSG--E 227

Qy        250 IMXAREAIALFGL--EPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTALT 307

Db        228 VLNAKEAFELASINSEFFELQPKEGLALVNGTAVGSGLASMVLFEANILAVLSEVLSAIF 287

Qy        308 VEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVH----HEEEVKVKDDEG 363

Db        288 AEVMQGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSYMKAAKKLHEIDPLQKP--- 343
```

```
Qy        364 ILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAG-QSTTDNPLIDVENKXTHHGGNFQA 422
Db        344 --KQDRYALRTSPQWLGPLIE--VIRFSTKSIEREINSVNDNPLIDVSRNKALHGGNFQG 399

Qy        423 SAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAA 481
Db        400 TPIGVSMDNTRLALASIGKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKGAEIAM 459

Qy        482 AAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQA 541
Db        460 ASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTNEAIEILKLMSSTFLIALCQA 519

Qy        542 VDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXELXXKVXKXLXKRL--EQTNSYD 599
Db        520 IDLRHLEENLKNSVKNTVSQVSKR---ILTTGVNGELHPSRFCEKDLLKVVDREYIFSYI 576

Qy        600 LEPRWHDAFSXATGTVVEXLSSXXXXXVSLAAVNA----------WKVAXAEKAIS--L 646
Db        577 DDP------CSATYPLMQKLRQ---VLVDHALVNAECEKDVNSSIFQKIAIFEEELKNLL 627

Qy        647 TRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSFVREELGVKARRGDVFLGKQEVTIGSN 704
Db        628 PKEVEGA-RAAYESGKAAIPNKIQECRSYPLYKFVREELGT-----GLLTGEKVRSPGEE 681

Qy        705 VSRIYEAIKSGRINXVLVKML 725
Db        682 FDKLFTAMCQGKIIDPLMECL 702
```

RESULT 13
S66343
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999
C;Accession: S66343
R;Fukasawa-Akada, T.; Kung, S.; Watson, J.C.
Plant Mol. Biol. 30, 711-722, 1996
A;Title: Phenylalanine ammonia-lyase gene structure, expression, and evolution in Nicoti
A;Reference number: S66343; MUID:96194455; PMID:8624404
A;Accession: S66343
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-715 <FUK>
A;Cross-references: EMBL:M84466; NID:g170349; PIDN:AAA34122.1; PID:g170350
C;Genetics:
A;Gene: PAL1
A;Introns: 133/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;201-203/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;202/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             28.0%;  Score 943;  DB 2;  Length 715;
  Best Local Similarity   36.4%;  Pred. No. 3.7e-66;
  Matches  253;  Conservative  107;  Mismatches  254;  Indels   82;  Gaps   18;

Qy         68 ELDGYSLTLGDVVGAA---RKGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTG 122
Db         53 KLGGESLTVAQVAAIAVRDKSANGVKVELSEEARAGVKASSDWVMDSMNKGTDSYGVTTG 112

Qy        123 FGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLENS--LPLEVVRGAMTIRV 180
Db        113 FGATSHRRTKNGGALQKELIRFLNAGV---------FGNGTETSHTLPHSATRAAMLVRI 163

Qy        181 NSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKV 240
Db        164 NTLLQGYSGIRFEILEAITKLINSNITPCLPLRGTITASGDLVPLSYIAGLLTGRPNSKA 223

Qy        241 HVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTAVSASMATLALHDAHMLSL 298
Db        224 -VGPNG--ETLNAEEAFRVAGVNGGFFELQPKEGLALVNGTAVGSGMASMVLFDSNILAV 280

Qy        299 LSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIRTLLEGSXFAVHHEEEVKV 358
Db        281 MSEVLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIMEHILDGSSY-VKAAQKLHE 338

Qy        359 KDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQSTTDNPLIDVENKXTHHGG 418
Db        339 MDPLQKPKQDRYALRTSPQWLGPQI-EVIRAATKMIEREINSVNDNPLIDVSRNKALHGG 397

Qy        419 NFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLPSCL-AAEDPSLSYHCKGL 477
Db        398 NFQGTPIGVSMDNARLALASIGKLMFAQFSELVNDYYNNGLPSNLTASRNPSLDYGFKGA 457

Qy        478 DIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYC 537
Db        458 EIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISARKTAEAVDILKLMSSTYLVA 517

Qy        538 VLQAVDLRAMEFEFKKQFXPXXXXLLXQHF-----GXXXTXXXXXXELXXKVXKX----- 587
Db        518 LCQAIDLRHLEENLKNAVKNTVSQVAKRTLTMGANGELHPARFCEKELLRIVDREYLFAY 577

Qy        588 ----------LXKRLEQT------NSYDLEPRWHDAFSXATGTVVEXLSSXXXXXVSLAA 631
Db        578 ADDPCSCNYPLMQKLRQVLVDHAMNNGESEKNVNSSIFQKIGAFED---------ELKA 627

Qy        632 VNAWKVAXAEKAISLTRXVRXXFWXAPSSSSPAL--XYLSPRTRVLYSFVREELGVKARR 689
Db        628 VLPKEVESARAAL--------------ESGNPAIPNRITECRSYPLYRFVRKELGT---- 669

Qy        690 GDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725
Db        670 -ELLTGEKVRSPGEECDKVFTAMCNGQIIDPMLECL 704
```

RESULT 14
T46172
phenylalanine ammonia-lyase (EC 4.3.1.5) 2 [similarity] - Arabidopsis thaliana
N;Alternate names: protein T4D2.190
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Mar-2000
C;Accession: T46172; S52991
R;Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23025
A;Accession: T46172
A;Molecule type: DNA
A;Residues: 1-717 <NYA>
A;Cross-references: EMBL:AL132958; NID:g6434245; PIDN:CAB64229.1; PID:g6630746
A;Experimental source: cultivar Columbia; BAC clone T4D2
R;Wanner, L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
Plant Mol. Biol. 27, 327-338, 1995
A;Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
A;Reference number: S52990; MUID:95195160; PMID:7888622
A;Accession: S52991
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-205,'H',207-234,'E',236-717 <WAN>
A;Cross-references: EMBL:L33678; NID:g497420; PIDN:AAC18871.1; PID:g497421
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
C;Genetics:
A;Gene: PAL2
A;Map position: 3
A;Introns: 135/2
A;Note: T4D2.190
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;203-205/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;204/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             27.9%;  Score 939.5;  DB 2;  Length 717;
  Best Local Similarity   34.2%;  Pred. No. 6.9e-66;
  Matches  251;  Conservative  122;  Mismatches  262;  Indels   99;  Gaps   20;

Qy         49 TQLDIVEXXLADPXT------------DDXXE----------LDGYSLTLGDVVGAAR 84
```

```
Db     15 TKVAVTTKTLADPLNWGLAADQMKGSHLDEVKKMVEEYRRPVVNLGGETLTIGQVAAIST 74

Qy     85 KGRXVRVXDSDEIRXKIDKSVEFLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALL 142

Db     75 VGGSVKVELAETSRAGVKASSDWVMESMNKGTDSYGVTTGFGATSHRRTKNGTALQTELI 134

Qy    143 EHQLCGVLPTSXDSFXLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFL 202

Db    135 RFLNAGIFGNTKETC-------HTLPQSATRAAMLVRVNTLLQGYSGIRFEILEAITSLL 187

Qy    203 NHGITPIVPLRGTISASGDLSPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGL 262

Db    188 NHNISPSLPLRGTITASGDLVPLSYIAGLLTGRPNSKA-TGPDG--ESLTAKEAFEKAGI 244

Qy    263 EP--VVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHP 320

Db    245 STGFFDLQPKEGLALVNGTAVGSGMASMVLFEANVQAVLAEVLSAIFAEVMSGKP-EFTD 303

Qy    321 FLHDVTRPHPTQIEVARNIRTLLEGSXF-----AVHHEEEVKVKDDEGILRQDRYPLRTS 375

Db    304 HLTHRLKHHPGQIEAAAIMEHILDGSSYMKLAQKVHEMDPLQKP------KQDRYALRTS 357

Qy    376 PQWLGPLVSDMIHAHAVLSLEAG-QSTTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRL 434

Db    358 PQWLGPQIE--VIRQATKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRL 415

Qy    435 ALALIGKLNFTQLTEMLNAGMNRGLPSCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLAN 493

Db    416 AIAAIGKLMFAQFSELVNDFYNNGLPSNLTASSNPSLDYGFKGAEIAMASYCSELQYLAN 475

Qy    494 PVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKK 553

Db    476 PVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLVGICQAVDLRHLEENLRQ 535

Qy    554 QFXPXXXXLLXQHFGXXXTXXXXXXELXXKV---------------XKXLXKRL--EQTN 596

Db    536 ------------------TVKNTVSQVAKKVLTTGINGELHPSRFCEKDLLKVVDREQVF 577

Qy    597 SYDLEP--RWHDAFSXATGTVVEXLSSXXXXXXVSLAAVNAWKVAXAEKAIS--LTRXVRX 652

Db    578 TYVDDPCSATYPLMQRLRQVIVDHALSNGETEKNAVTSIFQKIGAFEEELKAVLPKEVEA 637

Qy    653 XFWXAPSSSSPALXYLSP-RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEA 711

Db    638 ARAAYGNGTAPIPNRIKECRSYPLYRFVREELGTK-----LLTGEKVVSPGEEFDKVFTA 692

Qy    712 IKSGRINXVLVKML 725

Db    693 MCEGKLIDPLMDCL 706
```

RESULT 15
T03663
phenylalanine ammonia-lyase (EC 4.3.1.5) - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Jul-1999
C;Accession: T03663
R;Pellegrini, L.; Rohfritsch, O.; Fritig, B.; Legrand, M.
Plant Physiol. 106, 877-886, 1994
A;Title: Phenylalanine ammonia-lyase in tobacco. Molecular cloning and gene expression d
A;Reference number: Z14995; MUID:95125127; PMID:7824656
A;Accession: T03663
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-712 <PEL>
A;Cross-references: EMBL:X78269; NID:g633596; PIDN:CAA55075.1; PID:g633597
A;Experimental source: cultivar Samsun NN; tissue-type leaves after infection by tobacco
C;Genetics:
A;Gene: PAL
C;Function:
A;Description: catalyzes the deamination of L-phenylalanine to trans-cinnamic acid
A;Note: strongly induced during the hypersensitive reaction to TMV or to a fungal elicit
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;198-200/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;199/Modified site: dehydroalanine (Ser) #status predicted

```
  Query Match             27.9%;  Score 938;  DB 2;  Length 712;
  Best Local Similarity   35.2%;  Pred. No. 9e-66;
  Matches  253;  Conservative  115;  Mismatches  260;  Indels   90;  Gaps   18;

Qy     49 TQLDIVEXXLADPXTDDXXELDGYSLTLGDVVGAARK--GRXVRVXDSDEIRXKIDKSVE 106

Db     33 SHLDEVKKMVAE-FRKPVVKLGGETLTVAQVAAIAAKDNAKTVKVELSEGARAGVKASSD 91

Qy    107 FLRXQLXNSV--YGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPTSXDSFXLGRGLE 164

Db     92 WVMDSMSKGTDSYGVTTGFGATSHRRTKNGGALQKELIRFLNAGV---------FGNGTE 142

Qy    165 --NSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGDL 222

Db    143 SCHTLPQSGTRAAMLVRINTLLQGYSGIRFEILEAITKLLNHNVTPCLPLRGTITASGDL 202

Qy    223 SPLSYIAAAITGHPDSKVHVXHEGXEKIMXAREAIALFGLEP--VVLGPKEGLGLVNGTA 280

Db    203 VPLSYIAGLLTGRPNSKA-IGPNG--ETLNAEEAFRVAGVNSGFFELQPKEGLALVNGTA 259

Qy    281 VSASMATLALHDAHMLSLLSQALTALTVEAMVGHAGSFHPFLHDVTRPHPTQIEVARNIR 340

Db    260 VGSGLASMVLFDANILAVFSEVLSAIFAEVMNGKP-EFTDHLTHKLKHHPGQIEAAAIME 318

Qy    341 TLLEGSXFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDMIHAHAVLSLEAGQS 400

Db    319 HILDGSSY-VKAPQKLHETDPLQKPKQDRYALRTSPQWLGPQIEVIRSATKMIEREI-NS 376

Qy    401 TTDNPLIDVENKXTHHGGNFQASAVXNTMEKTRLALALIGKLNFTQLTEMLNAGMNRGLP 460

Db    377 VNDNPLIDVSRNKALHGGNFQGTPIGVSMDNARLALASIGKLMFAQFSELVNDYYNNGLP 436

Qy    461 SCL-AAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARR 519

Db    437 SNLTAGRNPSLDYGFKGSEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISARK 496

Qy    520 TAEANDVLSLLLATHLYCVLQAVDLRAMEFEFKKQFXPXXXXLLXQHFGXXXTXXXXXXE 579

Db    497 TAEAVDILKLMSSTYLVALCQAIDLRHLEENLRN-----------------------A 531

Qy    580 LXXKVXKXLXKRLEQTNSYDLEP-----------------RWHDAFSXATGTVVEXLSS 621

Db    532 VKNTVSQVAKRTLTMGANGELHPSRFCEKDLLRVVDREYVFRYADDACSANYPLMQKLRQ 591

Qy    622 XXXXXVSLAAVNAWKVAXA-----EKAISLTRXVRXXFWXAPSSSSPALXYLSP------ 670

Db    592 ---VLVDHALENGENEKNANSSIFQKILAFEGELKAVLPKEVESARISLENGNPAIANRI 648

Qy    671 ---RTRVLYSFVREELGVKARRGDVFLGKQEVTIGSNVSRIYEAIKSGRINXVLVKML 725

Db    649 KECRSYPLYRFVREELG-----AELLTGEKVRSPGEECDKVFTAMCNGQIIDSLLECL 701
```

Search completed: September 9, 2004, 10:08:48
Job time : 32.1203 secs